EXPRESS MAIL NO. EM555262883US

PATENT APPLICATION
IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

DETERMINING COMMON FUNCTIONAL ALLELES IN A POPULATION AND USES THEREFORE

Inventors:

Patricia D. Murphy, Slingerlands, New York, USA, citizen of the United States of America

Albert P. Halluin (Reg. No. 25,227)
HOWREY & SIMON
1299 Pennsylvania Avenue, N.W.
Washington, D.C. 20004
Telephone: (650) 463-8100

Attorney's Docket No. 5371.16.US01

DETERMINING COMMON FUNCTIONAL ALLELES IN A POPULATION AND USES THEREFORE

This application is a continuation-in-part of co-pending application number 08/905,772, filed August 4, 1997 which is a continuation in part of co-pending application number 08/798,691, filed February 12, 1997, which is a continuation in part of co-pending application number 08,598,591, filed on February 12, 1996, which issues as U.S. Patent No. 5,654,155 on August 5, 1997 and is also a continuation-in-part of co-pending application number U.S. Patent application 09/084,471, filed May 22, 1998, each of which is hereby incorporated by referenced in its entirety.

FIELD OF THE INVENTION

The invention relates to methods for identifying functional alleles commonly occurring in a population, for finding new functional alleles, for determining the relative frequencies at which such alleles, for genetic and pharmacogenetic applications of the methods and products produced thereby.

BACKGROUND OF THE INVENTION

An increasing number of genes which play a role in many different diseases are being identified. Detection of mutations in such genes is instrumental in determining susceptibility to or diagnosing these diseases. Some diseases, such as sickle cell disease, are known to be monomorphic; *i.e.*, the disease is generally caused by a single mutation present in the population. In such cases where one or only a few known mutations are responsible for the disease, methods for detecting the mutations are targeted to the site within the gene at which they are known to occur. However, the mutation responsible for such a monomorphic disease can only be established in the first instance if there exists an accurate reference sequence for the non-pathological state.

In many other cases individuals affected by a given disease display extensive allelic heterogeneity. For example, more than 125 mutations in the human BRCA1 gene have been reported (Breast Cancer Information Core world wide web site at http://www.nchgr.nih.gov/dir/lab_transfer/bic, which became publicly available on

November 1, 1995; Friend, S. et al., 1995, Nature Genetics 11:238). Mutations in the BRCA1 gene are thought to account for roughly 45% of inherited breast cancer and 80-90% of families with increased risk of early onset breast and ovarian cancer (Easton, 1993, et al., American Journal of Human Genetics 52: 678-701).

Other examples of genes for which the population displays extensive allelic heterogeneity and which have been implicated in disease include CFTR (cystic fibrosis), dystrophin (Duchenne muscular dystrophy, and Becker muscular dystrophy), and p53 (Li-Fraumeni syndrome).

Breast cancer is also an example of a disease in which, in addition to allelic heterogeneity, there is genetic heterogeneity. In addition to BRCA1, the BRCA2 and BRCA3 genes have been linked to breast cancer. Similarly, the NFI and NFII genes are involved in neurofibromatosis (types I and II, respectively). Furthermore, hereditary non-polyposis colorectal cancer (HNPCC) is a disease in which four genes, MSH2, MLH1, PMS1, and PMS2, have been implicated. It is yet another example of a disease in which there is both allelic and genetic heterogeneity of mutations. A cDNA sequence for MSH2 has been deposited in GenBank as Accession No. U03911; and a cDNA sequence for MLH1 has been deposited in GenBank as Accession No. U40978.

Additionally, disease or disease susceptibility also results from the interaction of more than one gene or the interaction of an environmental, chemical or biological influence on one or more genes. For example, measles virus infects many people; some are immune due to vaccination or previous infection, some are infected but asymptomatic, some become sick with a rash, some develop an encephalitis and some die. Genetic susceptibility and many other factors are involved in the outcome.

A common misconception in the field of molecular genetics is that for any given gene there exists a single "normal" or "wild-type" sequence. Often, research into such wild-type sequences ends once a single sequence associated with normal function is identified. For example, information in GenBank concerning the BRCA1 sequence represented by GenBank Accession No. U14680 does not indicate a basis for whether this sequence is representative of the population at large. Even when polymorphisms of the BRCA1 gene were identified, no analysis was provided of the arrangement of such

sequence variations in a given allele (*i.e.*, the haplotype) (Miki et al., 1994, Science 266: 66-71).

In the fields of plant and animal breeding, the "wild-type" may not be the desirable or may be one of several possibilities. For some domesticated plants and animals, the "wild-type" of any gene may not even be known. In the Brassica family, debate exists as to exactly what is a wild cabbage plant, much less which of the many genes or traits constitutes a "wild-type". By definition, a wild-type is not pathological but sometimes this definition seems inappropriate. For example, the MacIntosh apple is propagated asexually exclusively. An inability to reproduce naturally may be considered the result of pathological mutation(s) but is none the less the desired trait. In other situations, different strains of a plant are cross-breed where each set of genes from each parent strain may be considered "wild-type".

Identification of a mutation provides for early diagnosis which is essential for effective treatment of many diseases. However, in order to identify a mutation, it is necessary to have an accurate understanding of the proper reference sequences which encode the non-pathological functional gene products occurring in the population. Prior research efforts and publications have neither suggested nor taught a systematic approach to both identify a functional allele of a given gene and determine the relative frequency with which the allele occurs in the population.

Certain wild-type sequences of a gene may be otherwise indistinguishable from others except under certain circumstances. For example, a gene involved in resistance or susceptibility to a certain infectious agent is only recognized when the individual plant or animal is exposed to the infectious agent. Likewise chemical sensitivity may be a wild-type which is pathological under only certain circumstances which may never occur in the individual. Drought tolerance traits are significant only under environmental stress which may or may not occur. Therefore, the type of wild-type sequence is of importance.

SUMMARY OF THE INVENTION

It is an object of the invention to provide an integrated, systematic process for determining the functional allele profile for a given gene in a population. In accordance with the invention, a functional allele profile contains 1) the identity of the key functional

allele or alleles for a given gene in the population, including the "consensus" sequence, and 2) the relative frequency with which these functional alleles occur in the population. Thus, the functional allele profile includes the identification of the consensus normal sequence, *i.e.*, the most commonly occurring functional allele.

The present invention, therefore, provides a normal sequence which is the most likely sequence to be found in the majority of the normal population, the (i.e., "consensus normal DNA sequence"). A consensus normal allele sequence of a gene more accurately reflects the most likely sequence to be found in the population. Determining the consensus sequence is useful in both the diagnosis and treatment of disease. For example, use of the consensus normal gene sequence reduces the likelihood of misinterpreting a "sequence variation" found in the normal population with a pathologic "mutation" (*i.e.* causes disease in the individual or puts the individual at a high risk of developing the disease). A consensus normal DNA sequence makes it possible for true pathological mutations to be easily identified or differentiated from polymorphisms.

With large interest in mutation and polymorphism testing such as cancer predisposition testing, misinterpretation of sequence data is a particular concern. Individuals diagnosed with cancer want to know their prognosis and whether their disease is caused by a heritable genetic mutation. Likewise for other disease and traits and those who manage or manipulate these traits. Relatives of those with cancer who have not yet been diagnosed with the disease are also concerned whether they carry such a heritable mutation. Carrying such a mutation may increase risk of contracting the disease sufficiently to warrant an aggressive surveillance program. Accurate and efficient identification of mutations in genes linked to disease is crucial for widespread diagnostic screening for hereditary diseases.

In addition, the consensus sequence, or other sequences identified in the functional allele profile, allow for the selection of therapeutically optimal nucleotide sequences to be administered in gene therapy or gene replacement, or optimal amino acid sequence in the therapeutic administration of active proteins or peptides. The consensus sequence is generally the easiest target for various agonists, antagonists and measuring interactions with the gene or expression product appropriate for pharmacogenetic analysis.

Moreover, determining a functional allele profile of genes allows for an evaluation of the degree to which the gene is under selective pressure.

It is another embodiment of the present invention to find a new allele having a different wild-type haplotype from that previously known.

It is another embodiment of the present invention to determine the haplotype of a sample by determining the polymorphisms constituting the haplotype. Such a technique applies to one and plural genes, especially genes which interact or express products which interact with each other directly, interact with the same or similar other compound or are along the same metabolic pathway. As such, the method of the present invention determines combinations of haplotypes in different genes.

It is another embodiment of the present invention is determining how an individual will react to a particular chemical, environmental or biological influence. It is a premise of the present invention that different wild-type genes or their expression products interact differently in some circumstances.

Another embodiment of the present invention is the determination of traits and susceptibilities of plants and animals during breeding experiments by detecting the polymorphisms constituting the gene haplotype associated with the trait or susceptibility of interest.

BRIEF DESCRIPTION OF THE FIGURE

FIG. 1: Figure 1 shows alternative alleles containing polymorphic (non-mutation causing variations) sites along the BRCA1 gene, represented as individual "haplotypes" of the BRCA1 gene. The alternative allelic variations occurring at nucleotide positions 2201, 2430, 2731, 3232, 3667, 4427, and 4956 are shown. The BRCA1$^{(omi1)}$ haplotype is indicated with dark shading. For comparison, the haplotype available in GenBank is completely unshaded and designated as "GB". Two additional haplotypes (BRCA1$^{(omi2)}$, and BRCA1$^{(omi3)}$ are represented with mixed shaded and unshaded positions (numbers 7 and 9 from left to right).

DETAILED DESCRIPTION OF THE INVENTION

The invention provides an integrated, systematic process for determining the functional allele profile for a given gene or combination of genes in a population. In accordance with the invention, a functional allele profile contains 1) the identity of the key functional allele or alleles for a given gene in the population, including the "consensus" sequence, and 2) the relative frequency with which these functional alleles occur in the population. Thus, the functional allele profile includes the identification of the consensus normal sequence, *i.e.*, the most commonly occurring functional allele.

The present invention, therefore, provides a normal sequence which is the most likely sequence to be found in the majority of the normal population, the (i.e., "consensus normal DNA sequence"). A consensus normal allele sequence of a gene more accurately reflects the most likely sequence to be found in the population. In the process for determining functional alleles or afterward, one may search for and discover or synthesize a heretofor unknown or "new" allele.

A functional allele profile can be determined for any gene in which an altered or deficient function produces a recognizable, phenotypic trait, including, but not limited to, pathology. The invention is set forth for the purpose of illustration, and not by way of limitation, for determining the functional allele profile of three different genes associated with disease -- for example, the MSH2 and MLH1 genes, each associated with hereditary non-polyposis colorectal cancer (HNPCC), and the BRCA1 gene, associated with breast, ovarian, prostate and other cancers.

The following terms as used herein are defined as follows:

"Allele" refers to an alternative version (*i.e.,* nucleotide sequence) of a gene or DNA sequence at a specific chromosomal locus.

"Allelic variation" or "sequence variation" refers to a particular alternative nucleotide or nucleotide sequence at a position within a gene (*e.g.*, a polymorphic site or mutation) whose sequence varies from one allele to another.

"Coding sequence" or "DNA coding sequence" refers to those portions of a gene which, taken together, code for a peptide (protein), or which nucleic acid itself has function.

"Composite genomic sequence" refers to the combination of the two allelic nucleotide sequences (*i.e.*, maternal and paternal) obtained from sequencing a diploid genomic sample.

"Consensus" refers to the most commonly occurring in the population.

"Functional allele" refers to an allele which is naturally transcribed and translated into a functioning protein.

"Functional Allele Profile" refers to a set of functional alleles which are representative of the most common alleles occurring in a population, wherein the functional alleles are identified by nucleotide sequence and the relative frequencies with which the functional alleles occur in the population.

"Haplotype" refers to a set of nucleotides or nucleotide sequences occurring at sites of allelic variation occurring within a locus on a single chromosome (of either maternal or paternal origin). The "locus" includes the entire coding sequence.

"Mutation" refers to a base change or a gain or loss of base pair(s) in a DNA sequence, which results in a DNA sequence which codes for a non-functioning protein or a protein with substantially reduced or altered function.

"Agent for polymerization" refers to an enzyme which may be heat stable, e.g. *Taq* polymerase, or function at lower temperatures, e.g., room temperature, that effects an extension of DNA from a short primer sequence annealed to the target DNA of interest.

"Polymorphism" refers to an allelic variation which occurs in greater than or equal to 1% of the normal healthy population.

"Single nucleotide polymorphism" (SNP) refers to an allelic variation which is defined by two (and only two) alternative bases found at a specific and particular nucleotide in genomic DNA. It may be within a gene (i.e., exonic or intronic) or outside of a gene (such as in a promoter or other regulatory structure) or lastly found between genes.

"Individual" refers to a single organism which may be human, plant or non-human animal. The individual may be intact or a biological sample taken from the individual which contains sufficient substances or information regarding the individual.

"Protein variant" and "variant amino acid sequence" refers to different amino acid sequences from that in one naturally occurring wild-type protein and is generally

considered the same protein. Some different haplotypes have variant amino acid sequences.

"Expression product" refers to an RNA, spliced or unspliced, a pre-, pro-, prepro- or a peptide which alone or in conjunction with other peptides constitutes a protein.

"Pharmaceutical" refers to any bio-effecting chemical drug or biological agent which alters or induces an alteration in the metabolism of an "individual". Pharmaceuticals include compositions for use on veternary animals and agricultural and ornamental plants.

"Trait" refers to a phenotypically determinable characteristic resulting from the influence of one or more genes, alone or in conjunction with an environmental condition or exposure to other agents. Traits include susceptibilities to chemicals, infectious agents and environmental conditions (temperature, drought etc.).

Utility of the Invention

A person skilled in the art of genetic testing will find the present invention useful for diagnosis and treatment of diseases and susceptibility thereto. The invention is especially useful for establishing the "standard" (*i.e.*, consensus normal DNA sequence) and new haplotypes for clinical diagnostic, therapeutic, genetic testing and breeding uses.

Diagnostics

The diagnostic applications for which determining a functional allele profile in accordance with the invention include, but are not limited to, the following:

a) identifying individuals having a gene with no coding mutations, which individuals are therefore not at risk or have no increased susceptibility to the pathology(s) associated with a mutation in the gene in question;

b) avoiding misinterpretation of functional polymorphisms detected in the gene as mutations;

c) identifying individuals having a potentially abnormal gene that does not match the Consensus Normal DNA sequence;

d) determining ethnic founder haplotypes so that clinical analysis is appropriate for an individual from this ethnic group;

e) determining a sequence under strongest selective pressure; and

f) determining an amino acid and/or short nucleic acid sequence which may be derived from the consensus normal DNA sequence to make diagnostic and probes antibodies. Labeled diagnostic probes may be used by any hybridization method to determine the level of protein in serum or lysed cell suspension of a patient, or solid surface cell sample such as for immunohistochemical analysis.

g) detecting a new haplotype and determining the polymorphisms constituting the new haplotype.

h) detecting a new protein variant type and determining the variant amino acids constituting the new protein variant.

i) determining the combination of one haplotype or polymorphism for one gene and the haplotype or polymorphism for another different gene in the same individual. Generally, the genes or their expression products interact with each other directly, e.g. bind to each other, or indirectly by functioning with each other on the same substrate, are in different stages in a metabolic pathway, or are related to the same disease, susceptibility, condition or trait.

j) determining whether to administer a bioeffecting composition to an individual wherein individuals with different haplotypes for one or more genes respond differently to the composition.

k) determining susceptibility to disease or other pathology to decide on prophylaxis, therapy or differential monitoring.

l) determining a trait by quick assay of a genetic engineered or selectively bred individual. This permits one to determine the trait without actually measuring the trait phenotypically.

m) developing probe chips and panels of allele-specific oligonucleotide(s) to assay for the haplotypes or polymorphisms in one or more genes.

Therapeutics

Certain "normal" alleles may be more functional or hyper-functional than the minimum needed to maintain a normal phenotype in an individual, particularly when

stressed. By determining the most common allele in a population one may be observing empiric data for such suitability for survival (the effects may be so subtle that scientists have not determined the basis of this selection). For example, alleles with longer mRNA or protein half-lives (*i.e.*, stability) may produce healthier cells, and, thus, healthier people. Conversely, there may also be a selective advantage to a very short RNA half-life such as in proteins involved in the cell cycle pathway. Furthermore, proteases are known to have favored cutting sites which may be present or absent in different normal alleles leading to peptides that have intrinsic activity themselves.

Thus the determination of the functional allele profile or a new functional allele in accordance with the invention is useful in clinical therapy for:

a) selecting optimal alleles for performing gene repair or gene therapy; and

b) selecting optimal amino acid sequence for administration of functional protein in treatment or prevention of diseases.

Evolution and Population Genetics Analysis

The determination of the functional allele profile or a new functional allele in accordance with the invention is useful for:

a) determining whether a particular gene is under strong selective pressure; and

b) determining which of two or more genes which encode proteins with similar functions represents a redundant, or back-up copy of the gene.

Stepwise Process For Determining Functional Allele Profile

For the purpose of illustration, and not by way of limitation, the invention is described below for determining the functional allele profile of three cancer genes. However, the same principles can be applied in accordance with the invention to any gene in which a sequence variation results in a phenotypic trait, in any population within any species.

Screening for Individuals with Functional Allele Phenotype

In accordance with the invention, a group of individuals determined to be at low risk for carrying a mutation in the gene of interest is used as a source for genetic material.

Any standard method known in the art for performing pedigree analysis can be used for this selection process. See, for example, Harper, P.S., Practical Genetic Counseling, 3d. ed., 1988 (Wright/Butterworth & Co. Ltd.: Boston), especially at pages 4-7. For example, individuals can be screened in order to identify those with no disease history in their immediate family, *i.e.*, among their first and second degree relatives. A first degree relative is a parent, sibling, or offspring. A second degree relative is an aunt, uncle, grandparent, grandchild, niece, nephew, or half- sibling.

In a preferred embodiment for when a functional allele profile of an autosomal dominant disorder with relatively high penetrance (*e.g.*, greater than 50%) is desired, each person is asked to fill out a hereditary cancer prescreening questionnaire. More preferably, when an autosomal dominant cancer gene with such relatively high penetrance is the gene of interest, the questionnaire set forth in Table 1, below, is used.

Table 1
Hereditary Cancer Pre-Screening Questionnaire

Part A: Answer the following questions about your family

1. To your knowledge, has anyone in your family been diagnosed with a very specific hereditary colon disease called Familial Adenomatous Polyposis (FAP)?

2. To your knowledge, have you or any aunt had breast cancer diagnosed before the age 35?

3. Have you had Inflammatory Bowel Disease, also called Crohn's Disease or Ulcerative Colitis, for more than 7 years?

Part B: Refer to the list of cancers below for your responses **only** to questions in Part B

Bladder Cancer, Lung Cancer, Pancreatic Cancer, Breast Cancer, Gastric Cancer, Prostate Cancer, Colon Cancer, Malignant Melanoma, Renal Cancer, Endometrial Cancer, Ovarian Cancer, Thyroid Cancer

4. Have your mother or father, your sisters or brothers or your children had any of the listed cancers?

5. Have there been diagnosed in your mother's brothers or sisters, or your mother's parents more than one of the cancers in the above list?

6. Have there been diagnosed in your father's brothers or sisters, or your father's parents more than one of the cancers in the above list?

Part C: Refer to the list of relatives below for responses **only** to questions in Part C

You, Your mother, Your sisters or brothers, Your mothers's sisters or brothers (maternal aunts and uncles), Your children, Your mother's parents (maternal grandparents)

7. Have there been diagnosed in these relatives 2 or more identical types of cancer?

   Do not count "simple" skin cancer, also called basal cell or squamous cell skin cancer.

8. Is there a total of 4 or more of any cancers in the list of relatives above other than "simple" skin cancers?

Part D: Refer to the list of relatives below for responses **only** to questions in Part D.

You, Your father, Your sisters or brothers, Your fathers's sisters or brothers (paternal aunts and uncles)

Your children, Your father's parents (paternal grandparents)

9. Have there been diagnosed in these relatives 2 or more identical types of cancer?

   Do not count "simple" skin cancer, also called basal cell or squamous cell skin cancer.

10. Is there a total of 4 or more of any cancers in the list of relatives above other than "simple" skin cancers?

Individuals who answer no to all questions in Table 1 are designated as low risk of being carriers of mutations in the gene of interest and, therefore, in accordance with the invention, are candidates for further analysis set forth below.

Sequencing

From the group of individuals determined to have a low risk of being carriers for a mutant allele of the gene of interest, a group is selected for genomic DNA sequence analysis. Any number of samples may be analyzed. Preferably, a number of samples which is small enough for convenient, accurate sequence analysis, but large enough to provide a reliable representation of the population is analyzed. Most preferably, initial

sequencing may be performed on ten different chromosomes by analyzing samples from five unrelated individuals.

Preferably, sequencing template is obtained by amplifying the coding region and optionally one or more related sequences (e.g. splice site junctions, enhancers, introns, promotors and other regulatory elements) of the gene of interest. Any nucleic acid specimen, in purified or non-purified form, can be utilized as the starting nucleic acid or acids, providing it contains, or is suspected of containing, the specific nucleic acid sequence containing a polymorphic locus. Thus, the process may amplify, for example, DNA or RNA, including messenger RNA, wherein DNA or RNA may be single stranded or double stranded. In the event that RNA is to be used as a template, enzymes, and/or conditions optimal for reverse transcribing the template to DNA would be utilized. In addition, a DNA-RNA hybrid which contains one strand of each may be utilized. A mixture of nucleic acids may also be employed, or the nucleic acids produced in a previous amplification reaction herein, using the same or different primers may be so utilized. The specific nucleic acid sequence to be amplified, *i.e.,* the polymorphic locus, may be a fraction of a larger molecule or can be present initially as a discrete molecule, so that the specific sequence constitutes the entire nucleic acid. It is not necessary that the sequence to be amplified be present initially in a pure form; it may be a minor fraction of a complex mixture, such as contained in whole human DNA.

While the primer pairs used are greater than required to amplify the particular polymorphisms, the primer set actually used is listed below. For larger scale testing of polymorphisms for haplotype determination, only the primer pairs actually amplifying the polymorphism are required. Additionally, primers which amplify a shorter region, as short as the one nucleotide polymorphism may be used.

When a gene containing exons is analyzed, preferably the exonic sequences are individually amplified from genomic template DNA using a pair of primers specific for the intronic regions proximally bordering each individual exon.

DNA utilized herein may be extracted from a body sample, such as blood, tissue material and the like by a variety of techniques such as that described by Maniatis, *et. al.* in *Molecular Cloning:A Laboratory Manual,* Cold Spring Harbor, NY, pp. 280-281, 1982). If the extracted sample is impure, it may be treated before amplification with an

amount of a reagent effective to open the cells, or animal cell membranes of the sample, and to expose and/or separate the strand(s) of the nucleic acid(s). This lysing and nucleic acid denaturing step to expose and separate the strands will allow amplification to occur much more readily.

The deoxyribonucleotide triphosphates dATP, dCTP, dGTP, and dTTP are added to the synthesis mixture, either separately or together with the primers, in adequate amounts and the resulting solution is heated to about 90°-100°C from about 1 to 10 minutes, preferably from 1 to 4 minutes. After this heating period, the solution is allowed to cool, which is preferable for the primer hybridization. To the cooled mixture is added an appropriate agent for effecting the primer extension reaction (called herein "agent for polymerization"), and the reaction is allowed to occur under conditions known in the art. The agent for polymerization may also be added together with the other reagents if it is heat stable. This synthesis (or amplification) reaction may occur at room temperature up to a temperature above which the agent for polymerization no longer functions. Thus, for example, if DNA polymerase is used as the agent, the temperature is generally no greater than about 40°C. Most conveniently the reaction occurs at room temperature.

The primers used to carry out this invention embrace oligonucleotides of sufficient length and appropriate sequence to provide initiation of polymerization. Environmental conditions conducive to synthesis include the presence of nucleoside triphosphates and an agent for polymerization, such as DNA polymerase, and a suitable temperature and pH. The primer is preferably single stranded for maximum efficiency in amplification, but may be double stranded. If double stranded, the primer is first treated to separate its strands before being used to prepare extension products. The primer must be sufficiently long to prime the synthesis of extension products in the presence of the inducing agent for polymerization. The exact length of primer will depend on many factors, including temperature, buffer, and nucleotide composition. The oligonucleotide primer typically contains 12-20 or more nucleotides, although it may contain fewer nucleotides.

Primers used to carry out this invention are designed to be substantially complementary to each strand of the genomic locus to be amplified. This means that the primers must be sufficiently complementary to hybridize with their respective strands under conditions which allow the agent for polymerization to perform. In other words, the primers should have sufficient complementarity with the 5' and 3' sequences flanking the mutation to hybridize therewith and permit amplification of the genomic locus.

Oligonucleotide primers of the invention are employed in the amplification process which is an enzymatic chain reaction that produces exponential quantities of polymorphic locus relative to the number of reaction steps involved. Typically, one primer is complementary to the negative (-) strand of the polymorphic locus and the other is complementary to the positive (+) strand. Annealing the primers to denatured nucleic acid followed by extension with an enzyme, such as the large fragment of DNA polymerase I (Klenow) and nucleotides, results in newly synthesized + and - strands containing the target polymorphic locus sequence. Because these newly synthesized sequences are also templates, repeated cycles of denaturing, primer annealing, and extension results in exponential production of the region (*i.e.,* the target polymorphic locus sequence) defined by the primers. The product of the chain reaction is a discreet nucleic acid duplex with termini corresponding to the ends of the specific primers employed.

The oligonucleotide primers of the invention may be prepared using any suitable method, such as conventional phosphotriester and phosphodiester methods or automated embodiments thereof. In one such automated embodiment, diethylphosphoramidites are used as starting materials and may be synthesized as described by Beaucage, et al., Tetrahedron Letters, 22:1859-1862, (1981). One method for synthesizing oligonucleotides on a modified solid support is described in U.S. Patent No. 4,458,066.

The agent for polymerization may be any compound or system which will function to accomplish the synthesis of primer extension products, including enzymes. Suitable enzymes for this purpose include, for example, *E. coli* DNA polymerase I, Klenow fragment of *E. coli* DNA polymerase, polymerase muteins, reverse transcriptase, other enzymes, including heat-stable enzymes (*i.e.,* those enzymes which perform primer extension after being subjected to temperatures sufficiently elevated to cause

denaturation), such as *Taq* polymerase. Suitable enzyme will facilitate combination of the nucleotides in the proper manner to form the primer extension products which are complementary to each polymorphic locus nucleic acid strand. Generally, the synthesis will be initiated at the 3' end of each primer and proceed in the 5' direction along the template strand, until synthesis terminates, producing molecules of different lengths.

The newly synthesized strand and its complementary nucleic acid strand will form a double-stranded molecule under hybridizing conditions described above and this hybrid is used in subsequent steps of the process. In the next step, the newly synthesized double-stranded molecule is subjected to denaturing conditions using any of the procedures described above to provide single-stranded molecules.

The steps of denaturing, annealing, and extension product synthesis can be repeated as often as needed to amplify the target polymorphic locus nucleic acid sequence to the extent necessary for detection. The amount of the specific nucleic acid sequence produced will accumulate in an exponential fashion. Amplification is described in PCR. A Practical Approach, ILR Press, Eds. M. J. McPherson, P. Quirke, and G. R. Taylor, 1992.

The amplification products may be detected by Southern blots analysis, without using radioactive probes. In such a process, for example, a small sample of DNA containing a very low level of the nucleic acid sequence of the polymorphic locus is amplified, and analyzed via a Southern blotting technique or similarly, using dot blot analysis. The use of non-radioactive probes or labels is facilitated by the high level of the amplified signal. Alternatively, probes used to detect the amplified products can be directly or indirectly detectably labeled, for example, with a radioisotope, a fluorescent compound, a bioluminescent compound, a chemiluminescent compound, a metal chelator or an enzyme. Those of ordinary skill in the art will know of other suitable labels for binding to the probe, or will be able to ascertain such, using routine experimentation.

Sequences amplified by the methods of the invention can be further evaluated, detected, cloned, sequenced, and the like, either in solution or after binding to a solid support, by any method usually applied to the detection of a specific DNA sequence such as PCR, oligomer restriction (Saiki, *et al.*, Bio/Technology, 3:1008-1012, (1985)), allele-specific oligonucleotide (ASO) probe analysis (Conner, *et al.*, Proc. Natl. Acad.

Sci. U.S.A., 80:278, (1983)), oligonucleotide ligation assays (OLAs) (Landgren, *et al.*, Science, 241:1007, (1988)), heteroduplex analysis, chromatographic separation and the like. Molecular techniques for DNA analysis have been reviewed (Landgren, *et al.*, Science, 242:229-237, (1988)).

Preferably, the method of amplifying is by PCR, as described herein and as is commonly used by those of ordinary skill in the art. Alternative methods of amplification have been described and can also be employed as long as the genetic locus amplified by PCR using primers of the invention is similarly amplified by the alternative means. Such alternative amplification systems include but are not limited to self-sustained sequence replication, which begins with a short sequence of RNA of interest and a T7 promoter. Reverse transcriptase copies the RNA into cDNA and degrades the RNA, followed by reverse transcriptase polymerizing a second strand of DNA. Another nucleic acid amplification technique is nucleic acid sequence-based amplification (NASBA) which uses reverse transcription and T7 RNA polymerase and incorporates two primers to target its cycling scheme. NASBA can begin with either DNA or RNA and finish with either, and amplifies to 10^8 copies within 60 to 90 minutes. Alternatively, nucleic acid can be amplified by ligation activated transcription (LAT). LAT works from a single-stranded template with a single primer that is partially single-stranded and partially double-stranded. Amplification is initiated by ligating a cDNA to the promoter oligonucleotide and within a few hours, amplification is 10^8 to 10^9 fold. Another amplification system useful in the method of the invention is the QB Replicase System. The QB replicase system can be utilized by attaching an RNA sequence called MDV-1 to RNA complementary to a DNA sequence of interest. Upon mixing with a sample, the hybrid RNA finds its complement among the specimen's mRNAs and binds, activating the replicase to copy the tag-along sequence of interest. Another nucleic acid amplification technique, ligase chain reaction (LCR), works by using two differently labeled halves of a sequence of interest which are covalently bonded by ligase in the presence of the contiguous sequence in a sample, forming a new target. The repair chain reaction (RCR) nucleic acid amplification technique uses two complementary and target-specific oligonucleotide probe pairs, thermostable polymerase and ligase, and DNA nucleotides to geometrically amplify targeted sequences. A 2-base gap separates the

oligonucleotide probe pairs, and the RCR fills and joins the gap, mimicking DNA repair. Nucleic acid amplification by strand displacement activation (SDA) utilizes a short primer containing a recognition site for *HincII* with short overhang on the 5' end which binds to target DNA. A DNA polymerase fills in the part of the primer opposite the overhang with sulfur-containing adenine analogs. *HincII* is added but only cuts the unmodified DNA strand. A DNA polymerase that lacks 5' exonuclease activity enters at the cite of the nick and begins to polymerize, displacing the initial primer strand downstream and building a new one which serves as more primer. SDA produces greater than 10^7-fold amplification in 2 hours at 37°C. Unlike PCR and LCR, SDA does not require instrumented temperature cycling.

Another method is a process for amplifying nucleic acid sequences from a DNA or RNA template which may be purified or may exist in a mixture of nucleic acids. The resulting nucleic acid sequences may be exact copies of the template, or may be modified. The process has advantages over PCR in that it increases the fidelity of copying a specific nucleic acid sequence, and it allows one to more efficiently detect a particular point mutation in a single assay. A target nucleic acid is amplified enzymatically while avoiding strand displacement. Three primers are used. A first primer is complementary to the first end of the target. A second primer is complementary to the second end of the target. A third primer which is similar to the first end of the target and which is substantially complementary to at least a portion of the first primer such that when the third primer is hybridized to the first primer, the position of the third primer complementary to the base at the 5' end of the first primer contains a modification which substantially avoids strand displacement. This method is detailed in U.S. Patent 5,593,840 to Bhatnagar et al., 1997. Although PCR is the preferred method of amplification if the invention, these other methods can also be used to amplify the gene of interest.

A number of methods well-known in the art can be used to carry out the sequencing reactions. Preferably, enzymatic sequencing based on the Sanger dideoxy method is used. Mass spectroscopy may also be used.

The sequencing reactions can be analyzed using methods well-known in the art, such as polyacrylamide gel electrophoresis. In a preferred embodiment for efficiently

processing multiple samples, the sequencing reactions are carried out and analyzed using a fluorescent automated sequencing system such as the Applied Biosystems, Inc. ("ABI", Foster City, CA) system. For example, PCR products serving as templates are fluorescently labeled using the Taq Dye Terminator® Kit (Perkin-Elmer cat# 401628). Dideoxy DNA sequencing is performed in both forward and reverse directions on an ABI automated Model 377® sequencer. The resulting data can be analyzed using "Sequence Navigator®" software available through ABI.

Alternatively, large numbers of samples can be prepared for and analyzed by capillary electrophoresis, as described, for example, in Yeung et al., U.S. Patent No. 5,498,324.

Initial and Companion Haplotype Determination

The functional allele profiles identified in accordance with the invention may contain different alleles. Furthermore, each allele may contain multiple allelic variations, such as multiple polymorphisms. In other words, two different alleles may differ in sequence from one another at multiple nucleotide positions. Moreover, two such multiply polymorphic alleles may be present in the same individual, *i.e.*, a heterozygote. When the genomic sample of the gene of such a heterozygous individual is sequenced, the variations at each position can be detected. They are the alternative sequences present at particular positions in the composite sequence obtained from the diploid genome. However, at this stage, which variations are grouped together in each individual haplotype or allele, *i.e.*, the phase of the variations, cannot be determined.

For example, genomic sequence analysis of a hypothetical gene from a heterozygous individual may reveal that polymorphic positions 1, 2, or 3 each contain either an A or a G. However, it cannot be determined from this information alone whether the variations are distributed between the two alleles as:

allele 1 = $A_1A_2A_3$ and allele 2 = $G_1G_2G_3$; or

allele 1 = $A_1A_2G_3$ and allele 2 = $G_1G_2A_3$; or

allele 1 = $A_1G_2G_3$ and allele 2 = $G_1A_2A_3$, etc.

In accordance with the invention, such heterozygous genomic sequences obtained for the purpose of determining a functional allele profile are compared to an initial haplotype sequence. Some haplotypes can also be determined upon sequencing

chromosomal samples from a homozygous individual according to the methods above. Such homozygous sequence analyses contain no ambiguities in sequence between the two alleles because they are identical.

Preferably, an initial haplotype sequence is obtained by determining the cDNA sequence of an individual identified as being at low risk for carrying a mutation as described above. Because the full-length of a cDNA of the gene of interest is derived from a single mRNA transcript, it contains the allelic variations of a single haplotype. It contains all of the allelic variations present in a single allele of the individual from which it was obtained. Thus, the cDNA sequence contains half of the allelic variations present in the composite genomic sequence of a heterozygous individual containing that allele. Moreover, unlike sequence information from a heterozygous chromosomal sample, such cDNA sequence indicates which of the allelic variations are grouped together in one allele, *i.e.*, the phase of the variations.

By determining an initial haplotype, the companion haplotype present in a heterozygote can be determined by subtracting this sequence from the composite genomic sequence. For example, if in the illustration set forth above, the cDNA sequenced has an A in position 1, a G in position 2 and an A in position 3, then the initial haplotype is $A_1G_2A_3$. This sequence is then subtracted from the composite genomic sequence to yield the companion haplotype, namely $G_1A_2G_3$.

In general, the initial haplotype identified in a given individual also can be used to determine the presence of the haplotype in other individuals by comparing the initial haplotype sequence to the composite genomic sequence from such other individuals. When the number of allelic variations detected within a gene is four or greater, and especially when the number of allelic variations is five or greater, this method of subtracting the initial haplotype sequence from the composite genomic sequence of other individuals readily provides recognizably distinct haplotypes which are independent of each other. See, for example, the OMI^1 and GB haplotypes in FIG. 1, which differ from each other in each of seven sites of allelic variation.

When a haplotype determined in one individual is used to determine the haplotypes present in the composite genomic sequence of other individuals, the presence of that particular haplotype, and its companion haplotype as determined by subtraction

from a composite genomic sequence, should be confirmed. Such confirmation of the occurrence of a given haplotype in the population can be carried out, for example, by 1) sequencing cDNA samples, as described in this section, from such other heterozygous individuals; or 2) identifying individuals homozygous for the haplotype either among the initial set of sequenced chromosomal samples or by additional confirmatory sequencing of chromosomal samples as described below.

If an initial haplotype is not represented in any heterozygous composite genomic sequences obtained, one or more additional haplotypes should be obtained from such a heterozygous individual or from different individuals screened as above.

cDNA sequences for determining the initial haplotype can be obtained using standard techniques well known in the art. First, mRNA is isolated from an individual, for example, from blood or skin cells. The mRNA is initially reversed-transcribed into double stranded cDNA and then amplified according to the well known technique of RT-PCR (see, for example, U.S Patent No. 5,561,058 by Gelfand et al.).

The resulting cDNA, whose sequence represents a single haplotype, can be sequenced according to the methods above.

Determining the Relative Frequencies of the Haplotype

After all haplotypes have been identified in the study population, their relative frequencies are determined. For example, if five chromosomes out of a total of ten chromosomes are of one haplotype, then its frequency is 50%. Subsequently, each haplotype is ranked in order from the most frequent to the least frequent to yield the functional allele profile.

Confirmatory Analysis of Additional Samples

As described above, initial sequence analysis is performed on a small group of individuals, most preferably five individuals, screened according to the methods described above.

After identifying the haplotypes and determining their relative frequencies among the initial set of alleles analyzed, it may be desirable, in accordance with the invention, to perform follow-up, confirmatory sequencing on additional individuals who are also

screened according to the methods described above. Confirmatory sequencing can be carried out as above.

The haplotypes found occurring in the population are used as references to interpret the haplotypes present in any heterozygous individuals encountered during the confirmatory sequencing analysis of additional individuals.

By sequencing such additional samples, additional data points can be added to the functional allele profile to provide more precise frequencies of occurrence of each allele in the population. Furthermore, additional samples may contain a new functional allele with a new haplotype. This is particularly likely to be found for uncommon (<10%) or rare (<1%) haplotypes.

Furthermore, confirmatory sequence analysis ensures that the haplotypes determined by subtracting an initial haplotype from a composite heterozygous sequence is indeed represented in the population. Such techniques may also be used when multiple common haplotypes exist for the gene and it is uncertain which to use for subtraction.

When no sequence variation is found in the initial set of chromosomes, this indicates that the polymorphism rate of the gene of interest is uncommon (*e.g.*, polymorphisms occur in <10% of the alleles in the population studied). In such situations, identification of uncommon alleles and determination of their frequencies requires a confirmatory sequence analysis of samples from additional individuals. This method was used to detect such an uncommon polymorphism in exon 8 of the MLH1 gene, in Example 2 below.

Such confirmatory sequencing analysis also resulted in the identification and determination of relative frequency of occurrence of polymorphisms in intronic sequences, bordering exonic regions, of both the MSH2 and MLH1 genes, as detailed in Examples 1 and 2, respectively, below. The invention is illustrated by way of the Examples below.

EXAMPLE 1: Determining the Functional Allele Profile for MSH2

Approximately 150 volunteers are screened in order to identify individuals with no cancer history in their immediate family (i.e. first and second degree relatives). Each person is asked to fill out the hereditary cancer prescreening questionnaire shown in

Table 1, above. A first degree relative is a parent, sibling, or offspring. A second degree relative is an aunt, uncle, grandparent, grandchild, niece, nephew, or half-sibling. Among those individuals who answered "no" to all questions, five individuals are randomly chosen for end-to-end sequencing of their MSH2 gene.

Genomic DNA (100 nanograms) is extracted from white blood cells of five individuals designated as low risk of being carriers of mutations in the MSH2 gene from analysis of their answers to the questionnaire set forth in Table 1 above. The MSH2 coding region in each of the five samples is sequenced end-to-end by amplifying each exon individually. Each sample is amplified in a final volume of 25 microliters containing 1 microliter (100 nanograms) genomic DNA, 2.5 microliters 10X PCR buffer (100 mM Tris, pH 8.3, 500 mM KCl, 1.2 mM $MgCl_2$), 2.5 microliters 10X dNTP mix (2 mM each nucleotide), 2.5 microliters forward primer, 2.5 microliters reverse primer, and 1 microliter Taq polymerase (5 units), and 13 microliters of water.

The primers in Table 2, below, are used to carry out amplification of the various sections of the MSH2 gene samples. The primers are synthesized on an DNA/RNA Synthesizer Model 394®.

Table 2

MSH2 PRIMER SEQUENCES

Exon	Primer	Sequence
1	MSH1F-1	5'-CGC GTC TGC TTA TGA TTG G-3'
	MSH1R-1	5'-TCT CTG AGG CGG GAA AGG-3'
2	MSH2-2F-2-INSIDE	5'-TTT TTT TTT TTT TAA GGA GC-3'
	MSH2-2R-FULL	5'-CAC ATT TTT ATT TTT CTA CTC-3'
3	MSH3F	5'-GCT TAT AAA ATT TTA AAG TAT GTT C-3'
	MSH3R-2	5'-CTG GAA TCT CCT CTA TCA C-3'
4	MSH4F	5'-TTC ATT TTT GCT TTT CTT ATT CC-3'
	MSH4R	5'-ATA TGA CAG AAA TAT CCT TC-3'
5	MSH2-5F-1	5'-CAG TGG TAT AGA AAT CTT CGA-3'
	MSH2-5R-2-INSIDE	5'-TTT TTT TTT TTT TTA CCT GA-3'
6	MSH6F-1	5'-ACT AAT GAG CTT GCC ATT CT-3'
	MSH6R-1	5'-TGG GTA ACT GCA GGT TAC A-3'

7	MSH7F MSH7R	5'-GAC TTA CGT GCT TAG TTG-3' 5'-AGT ATA TAT TGT ATG AGT TGA AGG-3'
8	MSH8F MSH8R	5'-GAT TTG TAT TCT GTA AAA TGA GAT C-3' 5'-GGC CTT TGC TTT TTA AAA ATA AC-3'
9	MSH9F MSH9R	5'-GTC TTT ACC CAT TAT TTA TAG G-3' 5'-GTA TAG ACA AAA GAA TTA TTC C-3'
10	MSH10F MSH10R	5'-GGT AGT AGG TAT TTA TGG AAT AC-3' 5'CAT GTT AGA GCA TTT AGG G-3'
11	MSH11F MSH11R	5'-CAC ATT GCT TCT AGT ACA C-3' 5'-CCA GGT GAC ATT CAG AAC-3'
12	MSH12F MSH12R	5'-ATT CAG TAT TCC TGT GTA C-3' 5'-CGT TAC CCC CAC AAA GC-3'
13	MSH13F-1 MSH13R-1	5'ATG CTA TGT CAG TGT AAA CC-3' 5'CCA CAG GAA AAC AAC TAT TA-3'
14	MSH14F MSH14R	5'-TAC CAC ATT TTA TGT GAT GG-3' 5'-GGG GTA GTA AGT TTC CC-3'
15	MSH15F MSH15R	5'-CTC TTC TCA TGC TGT CCC-3' 5'-ATA GAG AAG CTA AGT TAA AC-3'
16	MSH16F MSH16R-1	5'-TAA TTA CTC ATG GGA CAT TC-3' 5'GGC ACT GAC AGT TAA CAC TA-3'

NOTE: These MSH2 primers are M-13 tailed:

M13 tail for F:5'-TGT AAA ACG ACG GCC AGT-3' added to 5' end of primer above

M13 tail for R:5'-CAG GAA ACA GCT ATG ACC-3' added to 5' end of primer above

Thirty-five cycles are performed, each consisting of denaturing (95°C; 30 seconds), annealing (55°C; 1 minute), and extension (72°C; 90 seconds), except during the first cycle in which the denaturing time was increased to 5 minutes, and during the last cycle in which the extension time was increased to 5 minutes.

PCR products are purified using Qia-quick® PCR purification kits (Qiagen®, cat# 28104; Chatsworth, CA). Yield and purity of the PCR product determined spectrophotometrically at OD_{260} on a Beckman DU 650 spectrophotometer.

All exons of the MSH2 gene are subjected to direct dideoxy sequence analysis by asymmetric amplification using the polymerase chain reaction (PCR) to generate a single stranded product amplified from this DNA sample. Shuldiner, *et al.,* Handbook of Techniques in Endocrine Research, p. 457-486, DePablo, F., Scanes, C., eds., Academic Press, Inc., 1993. Fluorescent dye is attached to PCR products for automated sequencing using the Taq Dye Terminator Kit (Perkin-Elmer® cat# 401628). DNA sequencing is performed in both forward and reverse directions on an Applied Biosystems, Inc. (ABI) Foster City, CA., automated sequencer (Model 377). The software used for analysis of the resulting data is "Sequence Navigator®" purchased through ABI.

Results

No differences in nucleotide sequence are observed among the coding exons of the five normal individuals (10 chromosomes), nor between these 10 chromosomal sequences and the sequence published in GenBank (Accession No. U03911) for MSH2. Thus, all ten individuals are homozygous for the same allele. An additional sixty-two normal individuals are sequenced end-to-end to confirm this result. Once again no sequence variation is found within the exons. However, minor variation in three single nucleotide polymorphisms are found in non-coding intronic sequences (IVS9-9; IVS10+6; IVS 10+12). The results are summarized in Table 3, below.

Table 3
MSH2 HAPLOTYPES

	Allelic Variations			
Haplotype	IVS9-9	IVS10+6	IVS10+12	Number of Chromosomes
GenBank sequence (U03911)	T	T	A	98 (73%)
Variant #1	A	C	G	28 (21%)
Variant #2	A	C	A*	6 (4.5%)
Variant #3	T	C**	A	2 (1.5%)

*Variant #2 is an uncommon derivative chromosome of variant #1

**Variant #3 is a rarer derivative chromosome of GenBank cDNA

Since the exonic coding sequence is maintained on all 4 haplotypes, such non-coding sequence variation did not result in any new "normal" coding consensus sequence of the MSH2 gene.

These results demonstrate that the sequence in the GenBank Repository is the "consensus normal DNA sequence" that should be used for comparison in all clinical applications to determine an individual with a hereditary susceptibility to HNPCC. In addition, these results indicate that normal MSH2 protein function, *i.e.*, mismatch repair function, is under a large degree of selective pressure to maintain viability in the human population. Very little if any variation in the activity of the MSH2 protein's mismatch repair function is tolerated, as reflected by the extraordinarily high degree of conservation of the normal sequence.

EXAMPLE 2: Determining the Functional Allele Profile for MLH1

All procedures (*e.g.*, selection of five individuals at low risk of being carriers for MLH1 mutations, isolation of genomic DNA, amplification of exons, sequencing of amplified exons, and analysis of sequence data) are carried out as described in Example 1, above, except that the amplification is carried out using primers specific to the MLH1 exons as set forth in Table 4, below.

Table 4

MLH1 PRIMER SEQUENCES

Exon	Primer	Sequence
1	MLHAF	5'-AGG CAC TGA GGT GAT TGG C-3'
	MLHAR	5'-TCG TAG CCC TTA AGT GAG C-3'
2	MLHBF-2	5'-TGA GGC ACT ATT GTT TGT ATT T-3'
	MLHBR-2	5'-TGT TGG TGT TGA ATT TTT CAG T-3'
3	MLHCF	5'-AGA GAT TTG GAA AAT GAG TAA C-3'
	MLHCR	5'-ACA ATG TCA TCA CAG GAG G-3'

4	MLHDF-1	5'-TGA GGT GAC AGT GGG TGA-3'
	MLHCR	5'-GAT TAC TCT GAG ACC TAG GC-3'
5	MLHEF	5'-GAT TTT CTC TTT TCC CCT TGG G-3'
	MLHER	5'-CAA ACA AAG CTT CAA CAA TTT AC-3'
6	MLHFF	5'-GGG TTT TAT TTT CAA GTA CTT CTA TG-3'
	MLHFR	5'-GCT CAG CAA CTG TTC AAT GTA TGA GC-3'
7	MLHGF	5'-CTA-GTG TGT GTT TTT GGC-3'
	MLHGR	5'-CAT AAC CTT ATC TCC ACC-3'
8	MLHHF	5'-CTC AGC CAT GAG ACA ATA AAT CC-3'
	MLHHR	5'-GGT TCC CAA ATA ATG TGA TGG-3'
9	MLHIF-1	5'-GTT TAT GGG AAG GAA CCT TGT-3'
	MLHIR-1	5'-TGG TCC CAT AAA ATT CCC TGT-3'
10	MLHJF	5'-CAT GAC TTT GTG TGA ATG TAC ACC-3'
	MLHJR	5'-GAG GAG AGC CTG ATA GAA CAT CTG-3'
11	MLHKF	5'-GGG CTT TTT CTC CCC CTC CC-3'
	MLHKR	5'-AAA ATC TGG GCT CTC ACG-3'
12	MLH1-LAF-2-INSIDE	5'-TTT AAT ACA GAC TTT GCT AC-3'
	MLH1-LBR	5'-GAA AAG CCA AAG TTA GAA GG-3'
13	MLHMF	5'-TGC AAC CCA CAA AAT TTG GC-3'
	MLHMR	5'-CTT TCT CCA TTT CCA AAA CC-3'
14	MLHNF	5'-TGG TGT CTC TAG TTC TGG-3'
	MLHNR	5'-CAT TGT TGT AGT AGC TCT GC-3'
15	MLHOF-2*	5'-GCA GAA CTA TGT CTG TCT CAT-3'
	MLHOR	5'-CGG TCA GTT GAA ATG TCA G-3'
16	MLHPF	5'-CAT TTG GAT CCG TTA AAG C-3'
	MLHPR	5'-CAC CCG GCT GGA AAT TTT ATT TG-3'
17	MLHQF	5'-GGA AAG GCA CTG GAG AAA TGG G-3'
	MLHQR	5'-CCC TCC AGC ACA CAT GCA TGT ACC G-3'
18	MLHRF	5'-TAA GTA GTC TGT GAT CTC CG-3'
	MLHRR	5'-ATG TAT GAG GTC CTG TCC-3'

19	MLHSF MLHSR*	5'-GAC ACC AGT GTA TGT TGG-3' 5'-GAG AAA GAA GAA CAC ATC CC-3'

NOTE: MLH1 primers are M-13 tailed,
*EXCEPT for MLH1 primers MLHOF-2, MLHOR & MLHSR:

M13 tail for F: 5'-TGT AAA ACG ACG GCC AGT-3' added to 5' end of primer above

M13 tail for R: 5'-CAG GAA ACA GCT ATG ACC-3' added to 5' end of primer above

Results

No differences are observed among the coding exons of the five normal individuals (10 chromosomes), nor between these 10 chromosomal sequences and the sequence published in GenBank (Accession No. U40978) for the MLH1 gene. In order to confirm these findings confirmatory sequencing is performed on an additional 62 samples. Among these sixty-two samples, variations are identified in only two positions as summarized in Table 5, below.

Table 5

MLH1 Haplotypes

	Allelic Variation		
Haplotype	EXON 8 codon 219	IVS14-19	Number of Chromosomes
GenBank Sequence (040978)	A	A	114 (92.5%)
Variant #1	A	G	5 (3.7%)
Variant #2	G	G	4 (3.1%)
Variant #3	G	A	1 (0.7%)
			Total 134 (100%)

One sequence variation is within exon 8 wherein a single nucleotide change from A to G in the first position of codon 219 (<u>A</u>TC --> <u>G</u>TC) changes the amino acid from Ile to Val. This sequence variation occurs approximately 3.7% of the time in this population.

The second sequence variation is deep within an intron (IV514-19) and can be found to be independently segregating with the exon 8 polymorphisms. While there were two "normal" exonic haplotypes identified in MLH1 (A versus G at codon 219), the most commonly found haplotype (*i.e.* consensus normal DNA sequence) having an A at the first position of codon 219 is the sequence currently in the GenBank database which should be used as the standard for clinical comparisons.

In addition, this analysis demonstrated that there is less selective pressure on the MLH1 gene (since codon 219 can have two forms) than on the MSH2 gene where no exonic sequence variation was tolerated. Given that these two genes are both mismatch repair genes, this observation indicates that the degree of redundancy of function (i.e., level of hierarchy between these proteins) is MSH2 as the primary system with MLH1 only as secondary or backup when MSH2 is dysfunctional (*i.e.*, mutant). While empiric data from other studies proposed such a relationship, only determining the actual functional allele profiles for these two genes provides an accurate understanding of the basis of previous observations from population studies.

EXAMPLE 3: Determining the Functional Allele Profile for BRCA1

All procedures (*e.g.*, selection of five individuals at low risk of being carriers for BRCA1 mutations, isolation of genomic DNA, amplification of exons, and sequencing of amplified exons, and analysis of sequence data) are carried out as described in Example 1, above, except that the amplification is carried out using primers specific to the BRCA1 exons as set forth in Table 6, below.

Table 6

BRCA1 PRIMERS FOR SEQUENCING TEMPLATES

Exon	Primer	SEQUENCE	Mg^{++}	SIZE
2	2F	5' GAAGTTGTCATTTTATAAACCTTT-3'	1.6	~275
	2R	5' TGTCTTTTCTTCCCTAGTATGT-3'		
3	3F	5' TCCTGACACAGCAGACATTA-3'	1.4	~375
	3R	5' TTGGATTTCGTTCTCACTTA-3'		
5	5F	5' CTCTTAAGGGCAGTTGTGAG-3'	1.2	~275
	5R	5' TTCCTACTGTGGTTGCTTCC-3'		
6	6/7F	5' CTTATTTTAGTGTCCTTAAAAGG-3'	1.6	~250
	6R	5' TTTCATGGACAGCACTTGAGTG-3'		

7	7F 6/7R	5' CACAACAAAGAGCATACATAGGG-3' 5' TCGGGTTCACTCTGTAGAAG-3'	1.6	~275
8	8F1 8R1	5' TTCTCTTCAGGAGGAAAAGCA-3' 5' GCTGCCTACCACAAATACAAA-3'	1.2	~270
9	9F 9R	5' CCACAGTAGATGCTCAGTAAA TA-3' 5' TAGGAAAATACCAGCTTCATAGA-3'	1.2	~250
10	10F 10R	5' TGGTCAGCTTTCTGTAATCG-3' 5' GTATCTACCCACTCTCTTCTTCAG-3'	1.6	~250
11A	11AF 11AR	5' CCACCTCCAAGGTGTATCA-3' 5' TGTTATGTTGGCTCCTTGCT-3'	1.2	372
11B	11BF1 11BR1	5' CACTAAAGACAGAATGAATCTA-3; 5' GAAGAACCAGAATATTCATCTA-3'	1.2	~400
11C	11CF1 11CR1	5' TGATGGGGAGTCTGAATCAA-3' 5' TCTGCTTTCTTGATAAAATCCT-3'	1.2	~400
11D	11DF1 11DR1	5' AGCGTCCCCTCACAAATAAA-3' 5' TCAAGCGCATGAATATGCCT-3'	1.2	~400
11E	11EF 11ER	5' GTATAAGCAATATGGAACTCGA-3' 5' TTAAGTTCACTGGTATTTGAACA-3'	1.2	388
11F	11FF 11FR	5' GACAGCGATACTTTCCCAGA-3' 5' TGGAACAACCATGAATTAGTC-3'	1.2	382
11G	11GF 11GR	5' GGAAGTTAGCACTCTAGGGA-3' 5' GCAGTGATATTAACTGTCTGTA-3'	1.2	423
11H	11HF 11HR	5' TGGGTCCTTAAAGAAACAAAGT-3' 5' TCAGGTGACATTGAATCTTCC-3'	1.2	366
11I	11IF 11IR	5' CCACTTTTTCCCATCAAGTCA-3' 5' TCAGGATGCTTACAATTACTTC-3'	1.2	377
11J	11JF 11JR	5' CAAAATTGAATGCTATGCTTAGA-3' 5' TCGGTAACCCTGAGCCAAAT-3'	1.2	377
11K	11KF 11KR-1	5' GCAAAAGCGTCCAGAAAGGA-3' 5' TATTTGCAGTCAAGTCTTCCAA-3'	1.2	396
11L	11LF-1 11LR	5' GTAATATTGGCAAAGGCATCT-3' 5' TAAAATGTGCTCCCCAAAAGCA-3'	1.2	360
12	12F 12R	5' GTCCTGCCAATGAGAAGAAA-3' 5' TGTCAGCAAACCTAAGAATGT-3'	1.2	~300
13	13F 13R	5' AATGGAAAGCTTCTCAAAGTA-3' 5' ATGTTGGAGCTAGGTCCTTAC-3'	1.2	~325
14	14F	5' CTAACCTGAATTATCACTATCA-3'	1.2	~310

	14R	5' GTGTATAAATGCCTGTATGCA-3'		
15	15F	5' TGGCTGCCCAGGAAGTATG-3'	1.2	~375
	15R	5' AACCAGAATATCTTTATGTAGGA-3'		
16	16F	5' AATTCTTAACAGAGACCAGAAC-3'	1.6	~550
	16R	5' AAAACTCTTTCCAGAATGTTGT-3'		
17	17F	5' GTGTAGAACGTGCAGGATTG-3'	1.2	~275
	17R	5' TCGCCTCATGTGGTTTTA-3'		
18	18F	5' GGCTCTTTAGCTTCTTAGGAC-3'	1.2	~350
	18R	5' GAGACCATTTTCCCAGCATC-3'		
19	19F	5' CTGTCATTCTTCCTGTGCTC-3'	1.2	~250
	19R	5' CATTGTTAAGGAAAGTGGTGC-3'		
20	20F	5' ATATGACGTGTCTGCTCCAC-3'	1.2	~425
	20R	5' GGGAATCCAAATTACACAGC-3'		
21	21F	5' AAGCTCTTCCTTTTTGAAAGTC-3'	1.6	~300
	21R	5' GTAGAGAAATAGAATAGCCTCT-3'		
22	22F	5' TCCCATTGAGAGGTCTTGCT-3'	1.6	~300
	22R	5' GAGAAGACTTCTGAGGCTAC-3'		
23	23F-1	5' TGAAGTGACAGTTCCAGTAGT-3'	1.2	~250
	23R-1	5' CATTTTAGCCATTCATTCAACAA-3'		
24	24F	5' ATGAATTGACACTAATCTCTGC-3'	1.4	~285
	24R	5' GTAGCCAGGACAGTAGAAGGA-3'		

[1] M13 tailed

Results

Differences in the nucleotide sequences of the five normal individuals are found in seven locations on the gene. The data show that for each of the samples, the BRCA1 gene is identical except in the region of seven single nucleotide polymorphisms. The changes and their positions are summarized on Table 7, below, and are depicted in schematic form in FIG. 1. The alternative alleles containing polymorphic (non-mutation causing allelic variations) sites along the BRCA1 gene are represented in FIG. 1 as individual "haplotypes" of the BRCA1 gene. The BRCA1$^{(omi1)}$ haplotype is shown in FIG. 1 and indicated with dark shading. The alternative allelic variations occurring at nucleotide positions 2201, 2430, 2731, 3232, 3667, 4427, and 4956 are shown. For comparison, the haplotype previously available in GenBank (as Accession No. U14680) is completely unshaded and designated "GB". As can be seen, the most common,

"consensus" haplotype occurs in five separate chromosomes labeled with the OMI symbol (haplotypes 1-5 from left to right). Two additional haplotypes (BRCA1$^{(omi2)}$, and BRCA1$^{(omi3)}$ are represented with mixed shaded and unshaded positions (numbers 7 and 9 from left to right). In total, 7 of the ten 10 haplotypes identified in the group of five individuals tested are not the haplotype available in GenBank.

The changes, their positions, and their frequencies among the five individuals (ten chromosomes) initially analyzed are summarized on Table 7, below.

Table 7

NORMAL PANEL TYPING

AMINO ACID CHANGE	EXON	1	2	3	4	5	FREQUENCY	
SER(SER) (694)	11E	C/C	C/T	C/T	T/T	T/T	0.4 C	0.6 T
LEU(LEU) (771)	11F	T/T	C/T	C/T	C/C	C/C	0.4 T	0.6 C
PRO(LEU) (871)	11G	C/T	C/T	C/T	T/T	T/T	0.3 C	0.7 T
GLU(GLY) (1038)	11I	A/A	A/G	A/G	G/G	G/G	0.4 A	0.6 G
LYS(ARG) (1183)	11J	A/A	A/G	A/G	G/G	G/G	0.4 A	0.6 G
SER(SER) (1436)	13	T/T	T/T	T/C	C/C	C/C	0.5 T	0.5 C
SER(GLY) (1613)	16	A/A	A/G	A/G	G/G	G/G	0.4 A	0.6 G

Note that there is no requirement to sequence the additional normal individuals available, as has been done for MSH2 (Example 1, above) and MLH1 (Example 2, above) to more accurately determine the frequencies of uncommon polymorphisms. A common

haplotype (the "consensus") is readily evident as different from the GenBank sequence (FIG. 1, "GB") in 50% of chromosomes and indeed is homozygous in two normal individuals.

Thus, the "consensus" sequence of the BRCA (omi[1]) should be used as the only true standard for clinical diagnostic analysis in order to avoid misinterpreting polymorphisms as pathologic mutations.

In the alternative, one could compare the test sequence against all four of the BRCA1 functional haplotypes.

Example 4: Pharmacogenetic Analysis of Sulfa Drug Sensitivity

The glucose-6-phosphate dehydrogenase gene is located on the X chromosome. Individuals with certain sequence variations in the G6PDH gene lead relatively normal lives unless they are exposed to certain chemicals found in fava beans, primaquine and sulfonamide antibiotics (sulfisoxazole, sulfamethoxazole, sulfathiazole, sulfacetamide, etc.). Upon administration of such compounds to the individual, severe reactions including hemolytic anemia occur in individuals having certain haplotype(s) of the G6PDH gene. These individuals are generally of African and Mediterranean heritage. Because these sequence variations are otherwise of little importance, they have been called both polymorphisms and mutations in the literature. For the purposes of this application, they are called mutations to distinguish them from clear polymorphisms. Genetic analysis in chimpanzees and various human populations indicate that the probable natural "wild-type" is found in individuals sensitive to sulfonamide antibiotics. Beutler et al, Blood 74: 2550-2555 (1989).

A number of apparently inconsequential single nucleotide polymorphisms (SNPs) in the G6PDH gene are known including at intron 5 (PvuII site), nucleotides 202 (Nla III site), 376 (Fok I site), 1311 and 1116 (Pst I sites). These constitute and define the haplotype. Missense mutations occur at amino acids 32, 48, 58, 68, 106, 126, 131, 156, 163, 165, 181, 182, 188, 198, 213, 216, 227, 282, 285, 291, 317, 323, 335, 342, 353, 363, 385, 386, 387, 393, 394, 398, 410, 439, 447, 454, 459, 463 and amino acid 35 deleted. Many mutations are restricted to certain haplotypes. Thus, haplotype determination provides an indication of whether the individual is sensitive to the drugs listed above.

Experimental

Blood is drawn from 30 individuals of African-American heritage with urinary tract infections having bacteria sensitive to sulfa antibiotics and for whom treatment with trimethoprim-sulfamethiazole is otherwise deemed appropriate. 1 mg of genomic DNA from individuals is isolated from peripheral blood lymphocytes and amplified by PCR using the primers listed in Hirono et al, Proc. Natl. Acad. Sci. USA 85:3951-3954 (1988) and Beutler et al, Human Genetics 87:462-464 (1990) according to the methods in Example 1 above. Amplified fragments are divided into five aliquots and four of which are cleaved by a restriction enzyme, either PvuII, Nla III, Fok I or Pst I, according to the manufacturer's (Stratagene and New England Biolabs) instructions. The digests are electrophoresed in a 4% agarose gel (NuSieve, FMC) with 10 ml of ethidium bromide (10 mg/ml) and the number of bands counted under ultraviolet light. The number of bands indicates the presence or absence of restriction enzyme cleavage and presence of a particular nucleotide at the polymorphic site.

An oligonucleotide probe for determining the polymorphic site at nucleotide 1311 is listed in Beutler et al, Human Genetics 87:462-464 (1990). The fifth aliquot is immobilized on a membrane and an ASO (allele specific oligonucleotide) hybridization assay is performed according to the method of Example 5 below. The presence or absence of the label indicating hybridization is considered indicative of the presence of a particular nucleotide at the polymorphic site.

Individuals having a haplotype, particularly the polymorphism at nucleotide 1116, indicative of very low likelihood of a G6PDH mutation sensitive to sulfamethiazole are given 160 mg trimethoprim with 800 mg sulfamethiazole (SEPTRA DS). Individuals having a haplotype or polymorphism indicative of a possible presence of a G6PDH mutation sensitive to sulfamethiazole are given a different antibiotic (varied with the patient) to which their infecting organism was susceptible.

Confirmatory sequencing of both alleles (60 chromosomes) of the coding region of the G6PDH gene is later performed by the techniques of Example 1 to determine the presence of a sensitizing mutation. The haplotype(s) associated with a mutation and those not associated with a mutation are recorded. A panel of oligonucleotides bound to a

membrane or other solid phase such as a DNA chip distinguishing the haplotypes and/or the common mutations also is to become part of the present invention.

<u>Example 5: Pharmacogenetic Analysis of BRCA1, BRCA2, PTEN, BAP1, BARD1 and hRAD51 Haplotypes and the Use of Tamoxifen to Prevent Breast Cancer</u>

While every step in carcinogenesis is not known, the BRCA1, BRCA2, PTEN, BAP1, BARD1 and hRAD51 proteins are either involved in breast, ovarian, prostate and other cancer susceptibility, in the metabolic pathway of or interact with such proteins. It was determined that the most common form of heriditary breast and ovarian cancer, the BRCA1 185delAG mutation, was found essentially exclusively in one haplotype, namely haplotype OMI1 as defined in Example 1, Fig. 1 and U.S. Patent 5,654,155. As such it was applicants hypothesis that the haplotypes of other related and similar genes alone or in certain combinations provide an indication of association with breast and other cancers associated with these genes, e.g. ovarian, pancreatic, prostate, colon, etc.

The various treatments and prophylactics useful against the disease are also believed to be related to the haplotypes. It is already known that certain mutant genes result in different presentations of cancers and different treatment. For example, BRCA1 mutations in the early part of the coding sequence generally form cancers at a younger age than mutations in the later part of the coding sequence. Likewise, breast cancer arising from BRCA2 mutations are typically more sensitive to radiation treatment than other breast cancers. Since some of these proteins actually bind to each other, different combinations of haplotypes may bind with different avidity to each other and operate slightly differently under certain circumstances. Likewise for proteins which act at separate reactions within the tumor-suppressing mechanisms.

<u>Experimental</u>

Blood samples are drawn from 47 women prescribed tamoxifen to prevent breast cancer or having had breast cancer to prevent reoccurrence of breast cancer. The DNA sequence for BRCA1 is determined in the regions of the single nucleotide polymorphic sites which constitute the haplotype use the primers according to U.S. Patent 5,654,155.

Those of BRCA2 are determined by using the primers of U.S. Patent application 09/084,471 filed May 22, 1998 or using the primers:

TABLE 8
BRCA2 PRIMERS

EXON	SEQUENCE	POLYMORPHISM
10AF	5'GAATAATATAAATTATATGGCTTA 3'	1093
10AR	5'CCTAGTCTTGCTAGTTCTT 3'	1093
10BF	5'ARCTGAAGTGGAACCAAATGATAC 3'	1593
10BR	5'ACGTGGCAAAGAATTCTCTGAAGTAA 3'	1593
11BF	5'AAGAAGCAAAATGTAATAAGGA 3'	2457
11BR	5'CATTTAAAGCACATACATCTTG 3'	2457
11CF	5'TCTAGAGGCAAAGAATCATAC 3'	2908
11CR	5'CAAGATTATTCCTTTCATTAGC 3'	2908
11DF	5'AACCAAAACACAAATCTAAGAG 3'	3199
11DR	5'GTCATTTTTATATGCTGCTTTAC 3'	3199
11EF	5'GGTTTTATATGGAGACACAGG 3'	3624
11ER	5'GTATTTACAATTTCAACACAAGC 3'	3624
11FF	5'ATCACAGTTTTGGAGGTAGC 3'	4035
11FR	5'CTGACTTCCTGATTCTTCTAA 3'	4035
14F	5'ACCATGTAGCAAATGAGGGTCT 3'	7470
14R	5'GCTTTTGTCTGTTTTCCTCCAA 3'	7470
22F	5'AACCACACCCTTAAGATGA 3'	9079
22R	5'GCATAAGTAGTGGATTTTGC 3'	9079

The DNA sequences for haplotypes of PTEN are determined by using the published primers of Table 3, Liaw et al, Nature Genetics, 16(1): p. 64-67 (1997).

The primers for amplifying hRAD51 are:
5'GGGCCCGGATCCATGGCAATGCAGATGCAGC 3' and
5'GGGCCCCAATGGATATCATTCAGTCTTTGGCATCTCCCACTCC 3'

The primers for amplifying BAP1 are:

PRIMER	SEQUENCE
BAP1A-F	5' CACGAGGCATGGCGCTGAGG 3'
BAP1A-R	5' CCGGGCCTTGTCTGTCCACT 3'
BAP1B-F	5' GTCTACCCCATTGACCATGG 3'
BAP1B-R	5' TCATCATCTGAGTACTGCTG 3'

BAP1C-F 5' TGCAGGAGGAAGAAGACCTG 3'

BAP1C-R 5' TCTGTCAGCGCCAGGGGACT 3'

BAP1D-F 5' AGCACAGGCCTGCTGCACCT 3'

BAP1D-R 5' GAAAAGGGGAAGTGGGGCAG 3'

The primers for amplifying BAP1 for polymorphism detection in the 3' UTR are:

BAP1-PF 5'AGCCCAGGCCCCAACACAGCCCCATGGCCTCT 3'

BAP1-PR 5'CTTAGGAGAGTTTATTCATTCATTGATCCAG 3'

The primers for amplifying BARD1 are:

5'AACAGTACAATGACTGGGCTC 3' and

5'TCAGCGCTTCTGCACACAGT 3'

In the cases of BARD1 and hRAD51, the PCR products are sequenced in entirety. All procedures (e.g., isolation of genomic DNA, amplification, sequencing, and analysis of sequence data) are carried out as described in Example 1. The method as described in Examples 1-3 is used to determine the common haplotypes in these genes.

Once standardized by sequencing, the amplified fragments of BRCA1, BRCA2, PTEN and BAP1, produced by PCR are assayed by hybridization to allele-specific oligonucleotides (ASO) which distinguish the polymorphic site directly. The ASO assay is performed as described in the following experiment.

Binding PCR Products to Nylon Membrane

The PCR products are denatured no more than 30 minutes prior to binding the PCR products to the nylon membrane. To denature the PCR products, the remaining PCR reaction (45 ml) and the appropriate positive control mutant gene amplification product are diluted to 200 ml final volume with PCR Diluent Solution (500 mM NaOH, 2.0 M NaCl, 25 mM EDTA) and mixed thoroughly. The mixture is heated to 95°C for 5 minutes, and immediately placed on ice and held on ice until loaded onto dot blotter, as described below.

The PCR products are bound to 9 cm by 13 cm nylon ZETA PROBE BLOTTING MEMBRANE (BIO-RAD, Hercules, CA, catalog number 162-0153) using a BIO-RAD dot blotter apparatus. Forceps and gloves are used at all times throughout the ASO analysis to manipulate the membrane, with care taken never to touch the surface of the membrane with bare hands or latex gloves.

Pieces of 3MM filter paper [WHATMAN®, Clifton, NJ] and nylon membrane are pre-wet in 10X SSC prepared fresh from 20X SSC buffer stock. The vacuum apparatus is rinsed thoroughly with dH_2O prior to assembly with the membrane. 100 ml of each denatured PCR product is added to the wells of the blotting apparatus. Each row of the blotting apparatus contains a set of reactions for a single exon to be tested, including a placental DNA (negative) control, a synthetic oligonucleotide with the desired mutation or a PCR product from a known mutant sample (positive control), and three no template DNA controls.

After applying PCR products, the nylon filter is placed DNA side up on a piece of 3MM filter paper saturated with denaturing solution (1.5M NaCl, 0.5 M NaOH) for 5 minutes. The membrane is transferred to a piece of 3MM filter paper saturated with neutralizing solution (1M Tris-HCl, pH 8, 1.5 M NaCl) for 5 minutes. The neutralized membrane is then transferred to a dry 3MM filter DNA side up, and exposed to ultraviolet light (STRALINKER, STRATAGENE, La Jolla, CA) for exactly 45 seconds to fix the DNA to the membrane. This UV crosslinking should be performed within 30 min. of the denaturation/neutralization steps. The nylon membrane is then cut into strips such that each strip contains a single row of blots of one set of reactions for a single exon.

Hybridizing Labeled Oligonucleotides to the Nylon Membrane

Prehybridization

The strip is prehybridized at 52°C incubation using the HYBAID® (SAVANT INSTRUMENTS, INC., Holbrook, NY) hybridization oven. 2X SSC (15 to 20 ml) is preheated to 52°C in a water bath. For each nylon strip, a single piece of nylon mesh cut slightly larger than the nylon membrane strip (approximately 1" x 5") is pre-wet with 2X SSC. Each single nylon membrane is removed from the prehybridization solution and placed on top of the nylon mesh. The membrane/mesh "sandwich" is then transferred onto a piece of Parafilm™. The membrane/mesh sandwich is rolled lengthwise and placed into an appropriate HYBAID® bottle, such that the rotary action of the HYBAID® apparatus caused the membrane to unroll. The bottle is capped and gently rolled to cause the membrane/mesh to unroll and to evenly distribute the 2X SSC, making sure that no air bubbles formed between the membrane and mesh or between the mesh

and the side of the bottle. The 2X SSC is discarded and replaced with 5 ml TMAC Hybridization Solution, which contained 3 M TMAC (tetramethyl ammoniumchloride - SIGMA T-3411), 100 mM Na_3PO_4(pH 6.8), 1 mM EDTA, 5X Denhardt's (1% Ficoll, 1% polyvinylpyrrolidone, 1% BSA (fraction V)), 0.6% SDS, and 100 mg/ml Herring Sperm DNA. The filter strips were prehybridized at 52°C with medium rotation (approx. 8.5 setting on the HYBAID® speed control) for at least one hour. Prehybridization can also be performed overnight.

Labeling Oligonucleotides

The DNA sequences of the oligonucleotide probes used to detect the BRCA1, BRCA2, PTEN, and BAP1 single nucleotide polymorphisms (SNPs) are as follows (for each polymorphism both options for the oligonucleotide are given below): The complements of these probes may also be used. Preliminary laboratory data indicates that probes with either greater specificity or sensitivity can be prepared by slightly varing the length and amount overlapping each side of the polymorphic region. It is expected that better probes will be prepared by routine experimentation.

TABLE 9 - BRCA1

2201 C 5' ACATGACAGCGATACTT 3'
2201 T 5' ACATGACAGTGATACTT 3'

2430 T 5' AGTATTTCATTGGTACC 3'
2430 C 5' AGTATTTCACTGGTACC 3'

2731 C 5' CATTTGCTCCGTTTTCA 3'
2731 T 5' CATTTGCTCTGTTTTCA 3'

3232 A 5' TTTTTAAAGAAGCCAGC 3'
3232 G 5' TTTTTAAAGGAGCCAGC 3'

3667 A 5' GCGTCCAGAAAGGAGAG 3'
3667 G 5' GCGTCCAGAGAGGAGAG 3'

4427 T 5' AAGTGACTCTTCTGCCC 3'
4427 C 5' AAGTGACTCCTCTGCCC 3'

4956 A 5' TGTGCCCAGAGTCCAGC 3'
4956 G 5' TGTGCCCAGGGTCCAGC 3'

1186 A 5' GGAATAAGCAGAAACTG 3'

1186 G5' GGAATAAGCGGAAACTG 3'

2196 G5' AAAAGACATGACAGCGA 3'
2196 A5' AAAAGACATAACAGCGA 3'

3238 G5' AAGAAGCCAGCTCAAGC 3'
3238 A5' AAGAAGCCAACTCAAGC 3'

2202 G5' CATGACAGTGATACTTT 3'
2202 A5' CATGACAGTAATACTTT 3'

TABLE 10 - BRCA2

PROBE SEQUENCE
1093 A5'TAGGACATTGGCATTGA 3'
1093 C5'TAGGACATGTGGCATTGA 3'

1342 A5'CTTCTGATTTGCTACATT 3'
1342 C5'CTTCTGATGTGCTACATT 3'

1593 A5'GGCTTCTCTGATTTTGGT 3'
1593 G5'GGCTTCTCGGATTTTGGT 3'

2457 T 5'TTTTGAATATTGTACTGG 3'
2457 C5'TTTTGAATGTTGTACTGG 3'

2908 G5'ATTAGCTACTTGGAAGAC 3'
2908 A5'ATTAGCTATTTGGAAGAC 3'

3199 A5'CCATTTGTTCATGTAATC 3'
3199 G5'CCATTTGTCCATGTAATC 3'

3624 A5'TAGCTTGGTTTTCTAAAC 3'
3624 G5'TAGCTTGGCTTTCTAAAC 3'

4035 T 5'ATTGAAACAACAGAATCA 3'
4035 C5'ATTGAAACGACAGAATCA 3'

7470 A5'TGAAAATGTGATTTAGTT 3'
7470 G5'TGAAAATGCGATTTAGTT 3'

9079 G5'TTCCATGGCCTTCCTAAT 3'
9079 A5'TTCCATGGTCTTCCTAAT 3'

TABLE 11 - PTEN

132 C 5'CTTGAAGGCGTATACAGG 3'

132 T 5'CTTGAAGG<u>T</u>GTATACAGG 3'

TABLE 12 - BAP1

+1102 5'ATGGCCTCT<u>A</u>CCAGATGGC 3'
+1102 5'ATGGCCTCT<u>C</u>CCAGATGGC 3'
+1102 5'ATGGCCTCT<u>G</u>CCAGATGGC 3'
+1102 5'ATGGCCTCT<u>T</u>CCAGATGGC 3'

+1116 5'CAGATGGCTTTG<u>A</u>AAAAGG 3'
+1116 5'CAGATGGCTTTG<u>C</u>AAAAGG 3'
+1116 5'CAGATGGCTTTG<u>G</u>AAAAGG 3'
+1116 5'CAGATGGCTTTG<u>T</u>AAAAGG 3'

+1131 5'GATCCAA<u>A</u>CAGGCCCCTTT 3'
+1131 5'GATCCAA<u>C</u>CAGGCCCCTTT 3'
+1131 5'GATCCAA<u>G</u>CAGGCCCCTTT 3'
+1131 5'GATCCAA<u>T</u>CAGGCCCCTTT 3'

+1233 5'CCCTGTAAA<u>A</u>ACTGGATCA 3'
+1233 5'CCCTGTAAA<u>C</u>ACTGGATCA 3'
+1233 5'CCCTGTAAA<u>G</u>ACTGGATCA 3'
+1233 5'CCCTGTAAA<u>T</u>ACTGGATCA 3'

Each labeling reaction contains 2-μl 5X Kinase buffer (or 1μl of 10X Kinase buffer), 5μl gamma-ATP ^{32}P (not more than one week old), 1μl T4 polynucleotide kinase, 3μl oligonucleotide (20 μM stock), sterile H_2O to 10 μl final volume if necessary. The reactions are incubated at 37°C for 30 minutes, then at 65°C for 10 minutes to heat inactivate the kinase. The kinase reaction is diluted with an equal volume (10μl) of sterile dH_20 (distilled water).

The oligonucleotides are purified on STE MICRO SELECT-D, G-25 spin columns (catalog no. 5303-356769), according to the manufacturer's instructions. The 20μl synthetic oligonucleotide eluate is diluted with 80 μl dH_20 (final volume = 100 μl). The amount of radioactivity in the oligonucleotide sample is determined by measuring the radioactive counts per minute (cpm). The total radioactivity must be at least 2 million cpm. For any samples containing less than 2 million total, the labeling reaction is repeated.

<u>Hybridization with Oligonucleotides</u>

Approximately 2-5 million counts of the labeled oligonucleotide probe is diluted into 5 ml of TMAC hybridization solution, containing 40 μl of 20 μM stock of unlabeled alternative polymorphism oligonucleotide. The probe mix is preheated to 52°C in the hybridization oven. The pre-hybridization solution is removed from each bottle and replaced with the probe mix. The filter is hybridized for 1 hour at 52°C with moderate agitation. Following hybridization, the probe mix is decanted into a storage tube and stored at -20°C. The filter is rinsed by adding approximately 20 ml of 2x SSC + 0.1% SDS at room temperature and rolling the capped bottle gently for approximately 30 seconds and pouring off the rinse. The filter is then washed with 2x SSC + 0.1% SDS at room temperature for 20 to 30 minutes, with shaking.

The membrane is removed from the wash and placed on a dry piece of 3MM WHATMAN filter paper then wrapped in one layer of plastic wrap, placed on the autoradiography film, and exposed for about five hours depending upon a survey meter indicating the level of radioactivity. The film is developed in an automatic Film processor.

Control Hybridization with Normal Oligonucleotides

The purpose of this step is to ensure that the PCR products are transferred efficiently to the nylon membrane.

Following hybridization with the bound oligonucleotide, as described above, each nylon membrane is washed in 2X SSC, 0.1% SDS for 20 minutes at 65°C to melt off the bound oligonucleotide probes. The nylon strips are then prehybridized together in 40 ml of TMAC hybridization solution for at least 1 hour at 52°C in a shaking water bath. 2-5 million counts of each of the normal labeled oligonucleotide probes plus 40 μl of 20μM stock of unlabeled normal oligonucleotide are added directly to the container containing the nylon membranes and the prehybridization solution. The filter and probes are hybridized at 52°C with shaking for at least 1 hour. Hybridization can be performed overnight, if necessary. The hybridization solution is poured off, and the nylon membrane is rinsed in 2X SSC, 0.1% SDS for 1 minute with gentle swirling by hand. The rinse is poured off and the membrane is washed in 2X SSC, 0.1% SDS at room temperature for 20 minutes with shaking.

The nylon membrane is removed and placed on a dry piece of 3MM WHATMAN filter paper. The nylon membrane is then wrapped in one layer of plastic wrap and placed on autoradiography film. The exposure is for at least 1 hour.

For each sample, adequate transfer to the membrane is indicated by a strong autoradiographic hybridization signal. For each sample, an absent or weak signal when hybridized with its normal oligonucleotide, indicates an unsuccessful transfer of PCR product, and it is a false negative. The ASO analysis must be repeated for any sample that did not successfully transfer to the nylon membrane.

The pattern of hybridization using the probes from the panel according to Tables 9-12 determine the haplotype of the patient sample when compared to the known haplotypes.

The degree of breast, ovarian and other cancer prevention with and without tamoxifen and the degree of prevention of reoccurrence of breast and ovarian cancer with and without tamoxifen are compared for patients grouped by BRCA1, BRCA2, PTEN, BAP1, BARD1, hRAD51 haplotype separately and in all possible combinations using various proprietary data mining techniques similar to the *Recognizer*™ methodology described in U.S. Patent 5,642,936. Appropriate recommendations regarding the use of tamoxifen for patients of different haplotypes are then be made for patients with and without a history of breast or ovarian cancer.

While this example is a *retrospective* study and thus unacceptable for proof of efficacy for the U.S. Food and Drug Administration, *prospective* studies are also part of the present invention. In a prospective study, the test individuals have their haplotypes determined for each pertinent gene prior to determining whether or not they will be accepted for the drug trial or initiate tamoxifen therapy.

Example 6: Pharmacogenetic Analysis of a p53 polymorphism and the Appropriateness of the Human Papiloma Virus Vaccine

Human papiloma virus (HPV) currently infects up to 40 million Americans with at least one of about 80 different strains. Many strains of the virus cause veneral warts, vulval, penile and perianal cancers. One strain in particular, HPV-16, is believed to be responsible for about half of all cases of cervical cancer. Three other strains are

responsible for another 35% of all cervical cancer cases with HPV-18 causing malignant tumors while HPV-6 and HPV-11 usually forming benign lesions. HPV vaccines are made by MedImmune, Inc. (Gaithersburg, Maryland) and Merck & Co. Clinical trials have already begun.

While applicant does not wish to be bound by any theory, it is believed that HPV may induce cancer by interacting with p53 in a manner which inhibits the action of p53 to prevent runaway cell growth. It has been known that HPV protein E6 inactivates only p53 proteins from some individuals and not other individuals. Medcalf et al, Oncogene, 8: 2847-2851 (1993). Therefore, determining the haplotype(s) of the p53 gene is believed to indicate who is susceptible to cervical cancer induced by HPV and is therefore a candidate for a HPV vaccine.

Previous commercial p53 gene testing of patient samples performed by Oncormed, Inc. (the owner of this application) involved various sequencing techniques and functional assays for prognostic testing on various tumor samples and susceptibility testing of genomic samples in patients with an inherited mutant p53 gene (Li-Fraumeni Syndrome). While apparent single nucleotide polymorphisms were noticed, such results were not reported as the samples are suspected to contain p53 mutations and do not originate from healthy individuals without a genetic history indicating inheritance of two functional p53 alleles.

Only polymorphisms in the coding region are analyzed because women having cervical cancers are believed to have a p53 protein which is "in-activatable" because the coding sequence for p53 is usually not mutated in cervical cancers. Vogelstein et al, Cell, 70: 523-526 (1992). Thus, the haplotypes were determined based on the single nucleotide polymorphisms at codon 21 (which may be either GAC or GAT), codon 36 (which may be either CCG or CCA), codon 47 (which may be either CCG or TCG), codon 72 (which may be either CGC or CCC) and codon 213 (which may be either CGA or CGG).

Experimental protocol

Blood samples are from 53 healthy individuals having a history of veneral warts or at risk from exposure to HPV. Exposure is defined as an individual having regular sexual contact with an infected individual without a barrier preventing transmission of

HPV. These individuals have either stage I (normal) or stage II (inflammation) PAP smears. Some of the individuals had been previously treated for veneral warts with one or more of the following treatments: podophyllin, trichloroacetic acid, cryosurgury, cauderization or interferon. Also, blood samples are from 12 patients with a history of cervical cancer as defined by a stage IV (carcinoma in-situ) or greater PAP smear result. Note that individuals having a stage III PAP smear (dysplasia) are not included in this study. White blood cells are collected and genomic DNA is extracted from the white blood cells according to well-known methods (Sambrook, et al., Molecular Cloning, A Laboratory Manual, 2nd Ed., 1989, Cold Spring Harbor Laboratory Press, at 9.16 - 9.19).

PCR Amplification for Sequencing

The genomic DNA is used as a template to amplify a DNA fragment encompassing the site of the mutation to be tested. The 25 ml PCR reaction contains the following components: 1 ml template (100 ng/ ml) DNA, 2.5 ml 10X PCR Buffer (PERKIN-ELMER), 1.5 ml dNTP (2 mM each dATP, dCTP, dGTP, dTTP), 1.5 ml Forward Primer (10 mM), 1.5 ml Reverse Primer (10 mM), 0.5 ml (2.5 U total) AMPLITAQ GOLD™ TAQ DNA POLYMERASE or AMPLITAQ® TAQ DNA POLYMERASE (PERKIN-ELMER), 1.0 to 5.0 ml (25 mM) $MgCl_2$ (depending on the primer) and distilled water (dH_2O) up to 25 ml. All reagents for each exon except the genomic DNA can be combined in a master mix and aliquoted into the reaction tubes as a pooled mixture. The primers are listed below.

NAME	SEQUENCE	LENGTH	INTRON
2F	5'-TCATGCTGGATCCCCACTTTTCCTCTTG-3'	28	31
2R	5'-GGTGGCCTGCCCTTCCAATGGATCCACT-3'	28	3
3F	5'-AATTCATGGGACTGACTTTCTGCTCTTGTC-3'	30	6
3R	5'-TCCAGGTCCCAGCCCAACCCTTGTCC-3'	26	4
4F	5'-GTCCTCTGACTGCTCTTTTCACCCATCTAC-3'	30	2
4R	5'-GGGATACGGCCAGGCATTGAAGTCTC-3'	26	29
5F	5'-CTTGTGCCCTGACTTTCAACTCTGTCTC-3'	28	16
5R	5'-TGGGCAACCAGCCCTGTCGTCTCTCCA-3'	27	15
6F	5'-CCAGGCCTCTGATTCCTCACTGATTGCTC-3'	29	4
6R	5'-GCCACTGACAACCACCCTTAACCCCTC-3'	27	29
7F	5'-GCCTCATCTTGGGCCTGTGTTATCTCC-3'	27	3

7R	5'-GGCCAGTGTGCAGGGTGGCAAGTGGCTC-3'	28	5
8F	5'-GTAGGACCTGATTTCCTTACTGCCTCTTGC-3'	30	23
8R	5'-ATAACTGCACCCTTGGTCTCCTCCACCGC-3'	29	20
9F	5'-CACTTTTATCACCTTTCCTTGCCTCTTTCC-3'	30	3
9R	5'-AACTTTCCACTTGATAAGAGGTCCCAAGAC-3'	30	7
10F	5'-ACTTACTTCTCCCCCTCCTCTGTTGCTGC-3'	29	2
10R	5'-ATGGAATCCTATGGCTTTCCAACCTAGGAAG-3'	31	39
11F	5'-CATCTCTCCTCCCTGCTTCTGTCTCCTAC-3'	29	2
11R	5'-CTGACGCACACCTATTGCAAGCAAGGGTTC-3'	30	80

The term "INTRON" refers to the location in the intron where the primer anneals.

Alternatively the primers for exons 2 and 3 may be amplified together with primers:

p53-2/3F 5'GAAGCGTCTCATGCTGGAT 3'

p53-2/3R 5'GGGGACTGTAGATGGGTGAA 3'

For each exon analyzed, the following control PCRs are set up:

(1) "Negative" DNA control (100 ng placental DNA (SIGMA CHEMICAL CO., St. Louis, MO)

(2) Three "no template" controls

PCR for all exons is performed using the following thermocycling conditions:

Temperature	Time	Number of Cycles
95°C	5 min. (AMPLITAQ)	1
or	10 min. (GOLD)	
95°C	30 sec. \	
55°C	30 sec. }	30 cycles
72°C	1 min /	
72°C	5 min.	1
4°C	hold	1

Quality control agarose gel of PCR amplification:

The quality of the PCR products is examined prior to further analysis by electrophoresing an aliquot of each PCR reaction sample on an agarose gel. 5 µl of each PCR reaction is run on an agarose gel along side a DNA 100 BP DNA LADDER (Gibco BRL cat# 15628-019). The electrophoresed PCR products are analyzed according to the following criteria:

Each patient sample must show a single band of the size corresponding the number of base pairs expected from the length of the PCR product from the forward primer to the reverse primer. If a patient sample demonstrates smearing or multiple bands, the PCR reaction must be repeated until a clean, single band is detected. If no PCR product is visible or if only a weak band is visible, but the control reactions with placental DNA template produced a robust band, the patient sample should be re-amplified with 2X as much template DNA.

All three "no template" reactions must show no amplification products. Any PCR product present in these reactions is the result of contamination. If any one of the "no template" reactions shows contamination, all PCR products should be discarded and the entire PCR set of reactions should be repeated after the appropriate PCR decontamination procedures have been taken.

The optimum amount of PCR product on the gel should be between 50 and 100 ng, which can be determined by comparing the intensity of the patient sample PCR products with that of the DNA ladder. If the patient sample PCR products contain less than 50 to 100 ng, the PCR reaction should be repeated until sufficient quantity is obtained.

DNA Sequencing

For DNA sequencing, double stranded PCR products are labeled with four different fluorescent dyes, one specific for each nucleotide, in a cycle sequencing reaction. With Dye Terminator Chemistry, when one of these nucleotides is incorporated into the elongating sequence it causes a termination at that point. Over the course of the cycle sequencing reaction, the dye-labeled nucleotides are incorporated along the length of the PCR product generating many different length fragments.

The dye-labeled PCR products will separate according to size when electrophoresed through a polyacrylamide gel. At the lower portion of the gel on an ABI

automated sequencer, the fragments pass through a region where a laser beam continuously scans across the gel. The laser excites the fluorescent dyes attached to the fragments causing the emission of light at a specific wavelength for each dye. Either a photomultiplier tube (PMT) detects the fluorescent light and converts is into an electrical signal (ABI 373) or the light is collected and separated according to wavelength by a spectrograph onto a cooled, charge coupled device (CCD) camera (ABI 377). In either case the data collection software will collect the signals and store them for subsequent sequence analysis.

PCR products are first purified for sequencing using a QIAQUICK-SPIN PCR PURIFICATION KIT (QIAGEN #28104). The purified PCR products are labeled by adding primers, fluorescently tagged dNTPs and Taq Polymerase FS in an ABI Prism Dye Terminator Cycle Sequencing Kit (PERKIN ELMER/ABI catalog #02154) in a PERKIN ELMER GENEAMP 9600 thermocycler.

The amounts of each component are:

For Samples		For Controls	
Reagent	Volume	Reagent	Volume
Dye mix	8.0 μL	PGEM	2.0 μL
Primer (1.6 mM)	2.0 μL	M13	2.0 μL
PCR product	2.0 μL	Dye mix	8.0 μL
sdH2O	8.0 μL	sdH2O	8.0 μL

The thermocycling conditions are:

Temperature	Time	# of Cycles
96°C	15 sec. \	
50°C	5 sec. }	25
60°C	4 min. /	
4°C	hold	1

The product is then loaded into a gel and placed into an ABI DNA Sequencer (Models 373A & 377) and run. The sequence obtained is analyzed by comparison to the wild type (reference) sequence using SEQUENCE NAVIGATOR software. When a sequence does not align, it indicates a possible mutation or polymorphism. The DNA

sequence is determined in both the forward and reverse directions. All results are provided to a second reader for review.

PCR Amplification for ASO

The genomic DNA is used as a template to amplify a separate DNA fragment encompassing the site of the mutation to be tested. The 50 µl PCR reaction contains the following components: 1 µl template (100 ng/ µl) DNA, 5.0 µl 10X PCR Buffer (PERKIN-ELMER), 2.5 µl dNTP (2mM each dATP, dCTP, dGTP, dTTP), 2.5 µl Forward Primer (10 mM), 2.5 µl Reverse Primer (10 µM), 0.5 µl (2.5 U total) AMPLITAQ® TAQ DNA POLYMERASE or AMPLITAQ GOLD™ DNA POLYMERASE (PERKIN-ELMER), 1.0 to 5.0 µl (25 mM) $MgCl_2$ (depending on the primer) and distilled water (dH_2O) up to 50 µl. All reagents for each exon except the genomic DNA can be combined in a master mix and aliquoted into the reaction tubes as a pooled mixture. The primers described above are used.

For each exon analyzed, the following control PCRs are set up:

(1) "Negative" DNA control (100 ng placental DNA (SIGMA CHEMICAL CO., St. Louis, MO)

(2) Three "no template" controls.

PCR for all exons is performed using the following thermocycling conditions:

Temperature		Time		Number of Cycles
95°C		5 min.(AMPLITAQ)		1
	or	10 min. (GOLD)		
95°C		30 sec.	\	
55°C		30 sec.	}	30 cycles
72°C		1 min	/	
72°C		5 min.		1
4°C		hold		1

The quality control agarose gel of PCR amplification is performed as above.

Binding PCR Products to Nylon Membrane

The PCR products are denatured no more than 30 minutes prior to binding the PCR products to the nylon membrane. To denature the PCR products, the remaining PCR reaction (45 µl) and the appropriate positive control polymorphism gene

amplification product are diluted to 200 µl final volume with PCR Diluent Solution (500 mM NaOH, 2.0 M NaCl, 25 mM EDTA) and mixed thoroughly. The mixture is heated to 95°C for 5 minutes, and immediately placed on ice and held on ice until loaded onto dot blotter, as described below.

The PCR products are bound to 9 cm by 13 cm nylon ZETA PROBE BLOTTING MEMBRANE (BIO-RAD, Hercules, CA, catalog number 162-0153) using a BIO-RAD dot blotter apparatus.

Pieces of 3MM filter paper [WHATMAN®, Clifton, NJ] and nylon membrane are pre-wet in 10X SSC prepared fresh from 20X SSC buffer stock. The vacuum apparatus is rinsed thoroughly with dH_2O prior to assembly with the membrane. 100 µl of each denatured PCR product is added to the wells of the blotting apparatus. Each row of the blotting apparatus contains a set of reactions for a single exon to be tested, including a placental DNA (negative) control, a synthetic oligonucleotide with the desired mutation or a PCR product from a known polymorphic sample (positive control), and three no template DNA controls.

After applying PCR products, the nylon filter is placed DNA side up on a piece of 3MM filter paper saturated with denaturing solution (1.5 M NaCl, 0.5 M NaOH) for 5 minutes. The membrane is transferred to a piece of 3MM filter paper saturated with neutralizing solution (1 M Tris-HCl, pH 8, 1.5 M NaCl) for 5 minutes. The neutralized membrane is then transferred to a dry 3MM filter DNA side up, and exposed to ultraviolet light (STRALINKER, STRATAGENE, La Jolla, CA) for exactly 45 seconds to fix the DNA to the membrane. This UV crosslinking should be performed within 30 min. of the denaturation/neutralization steps. The nylon membrane is then cut into strips such that each strip contains a single row of blots of one set of reactions for a single exon.

Hybridizing Labeled Oligonucleotides to the Nylon Membrane

Prehybridization

The strip is prehybridized at 52°C incubation using the HYBAID® (SAVANT INSTRUMENTS, INC., Holbrook, NY) hybridization oven. 2X SSC (15 to 20 ml) is preheated to 52°C in a water bath. For each nylon strip, a single piece of nylon mesh cut slightly larger than the nylon membrane strip (approximately 1" x 5") is pre-wet with 2X SSC. Each single nylon membrane is removed from the prehybridization solution and

placed on top of the nylon mesh. The membrane/mesh "sandwich" is then transferred onto a piece of Parafilm™. The membrane/mesh sandwich is rolled lengthwise and placed into an appropriate HYBAID® bottle, such that the rotary action of the HYBAID® apparatus caused the membrane to unroll. The bottle is capped and gently rolled to cause the membrane/mesh to unroll and to evenly distribute the 2X SSC, making sure that no air bubbles formed between the membrane and mesh or between the mesh and the side of the bottle. The 2X SSC is discarded and replaced with 5 ml TMAC Hybridization Solution, which contains 3 M TMAC (tetramethyl ammoniumchloride - SIGMA T-3411), 100 mM Na_3PO_4(pH 6.8), 1 mM EDTA, 5X Denhardt's (1% Ficoll, 1% polyvinylpyrrolidone, 1% BSA (fraction V)), 0.6% SDS, and 100 mg/ml Herring Sperm DNA. The filter strips are prehybridized at 52°C with medium rotation (approx. 8.5 setting on the HYBAID® speed control) for at least one hour. Prehybridization can also be performed overnight.

Labeling Oligonucleotides

The DNA sequences of the numerous oligonucleotide probes are used to detect the p53 mutation. For each mutation, a polymorphic and a normal oligonucleotide must be labeled. While only five pairs of oligonucleotide probes are listed below, corresponding oligonucleotides for each mutation may be prepared and used in the same manner.

Polymorphism in codon 21

wild-type 5'TTTTCAGACCTATGGAAAC 3'

other wt 5'TTTTCAGATCTATGGAAAC 3'

Polymorphism in codon 36

wild-type 5'CCCTTGCCGTCCCAAGCA 3'

other wt 5'CCCTTGCCATCCCAAGCA 3'

Polymorphism in codon 47

wild-type 5'CTGTCCCCGGACGATATT 3'

other wt 5'CTGTCCCCAGACGATATT 3'

Polymorphism in codon 72

wild-type 5'GCTCCCCCCGTGGCCCCT 3'

other wt	5'GCTCCCC<u>G</u>CGTGGCCCCT 3'	
		Polymorphism in codon 213
wild-type	5'ACTTTTCG<u>A</u>CATAGTGTG 3'	
other wt	5'ACTTTTCG<u>G</u>CATAGTGTG 3'	

Each labeling reaction contains 2 μl 5X Kinase buffer (or 1 μl of 10X Kinase buffer), 5 μl gamma-ATP ^{32}P (not more than one week old), 1 μl T4 polynucleotide kinase, 3 μl oligonucleotide (20 μM stock), sterile H_2O to 10 μl final volume if necessary. The reactions are incubated at 37°C for 30 minutes, then at 65°C for 10 minutes to heat inactivate the kinase. The kinase reaction is diluted with an equal volume (10 μl) of sterile dH_20 (distilled water).

The oligonucleotides are purified on STE MICRO SELECT-D, G-25 spin columns (catalog no. 5303-356769), according to the manufacturer's instructions. The 20 μl synthetic oligonucleotide eluate is diluted with 80 μl dH_20 (final volume = 100 μl). The amount of radioactivity in the oligonucleotide sample is determined by measuring the radioactive counts per minute (cpm). The total radioactivity must be at least 2 million cpm. For any samples containing less than 2 million cpm total, the labeling reaction is repeated.

<u>Hybridization with Oligonucleotides</u>

Approximately 2-5 million cpm of the labeled polymorphic oligonucleotide probe is diluted into 5 ml of TMAC hybridization solution, containing 40 μl of 20 μM stock of unlabeled normal oligonucleotide. The probe mix is preheated to 52°C in the hybridization oven. The pre-hybridization solution is removed from each bottle and replaced with the probe mix. The filter is hybridized for 1 hour at 52°C with moderate agitation. Following hybridization, the probe mix is decanted into a storage tube and stored at -20°C. The filter is rinsed by adding approximately 20 ml of 2x SSC + 0.1% SDS at room temperature and rolling the capped bottle gently for approximately 30 seconds and pouring off the rinse. The filter is then washed with 2x SSC + 0.1% SDS at room temperature for 20 to 30 minutes, with shaking.

The membrane is removed from the wash and placed on a dry piece of 3MM WHATMAN filter paper then wrapped in one layer of plastic wrap, placed on the autoradiography film, and exposed for about five hours depending upon a survey meter

indicating the level of radioactivity. The film is developed in an automatic film processor.

Control Hybridization with Normal Oligonucleotides

The purpose of this step is to ensure that the PCR products are transferred efficiently to the nylon membrane.

Following hybridization with the polymorphic oligonucleotide each nylon membrane is washed in 2X SSC, 0.1% SDS for 20 minutes at 65°C to melt off the polymorphic oligonucleotide probes. The nylon strips are then prehybridized together in 40 ml of TMAC hybridization solution for at least 1 hour at 52°C in a shaking water bath. 2-5 million counts of each of the normal labeled oligonucleotide probes plus 40 ml of 20 mM stock of unlabeled normal oligonucleotide are added directly to the container containing the nylon membranes and the prehybridization solution. The filter and probes are hybridized at 52°C with shaking for at least 1 hour. Hybridization can be performed overnight, if necessary. The hybridization solution is poured off, and the nylon membrane is rinsed in 2X SSC, 0.1% SDS for 1 minute with gentle swirling by hand. The rinse is poured off and the membrane is washed in 2X SSC, 0.1% SDS at room temperature for 20 minutes with shaking.

The nylon membrane is removed placed on a dry piece of 3MM WHATMAN filter paper. The nylon membrane is then wrapped in one layer of plastic wrap and placed on autoradiography film, and exposure is for at least 1 hour.

For each sample, adequate transfer to the membrane is indicated by a strong autoradiographic hybridization signal. For each sample, an absent or weak signal when hybridized with its normal oligonucleotide, indicates an unsuccessful transfer of PCR product, and it is a false negative. The ASO analysis must be repeated for any sample that did not successfully transfer to the nylon membrane.

Homozygous individuals having haplotypes with the single nucleotide polymorphism (SNP) arginine at codon 72 are overrepresentated in the genomic alleles of cervical cancer patients. In addition, it was recently published that cervical tumors have the SNP arginine at codon 72 at significantly higher frequency than normal tissue. Storey

et al, Nature, 393: 229-234 (1998). Healthy women having such haplotypes are candidates for the HPV vaccines to prevent HPV invection, treat veneral warts, treat cervical and other related cancers, and prevent reoccurrence of veneral warts previously treated.

Example 7: Pharmacogenetic Analysis of P1 Haplotype and Platelet Sensitivity to Aspirin

Aspirin has been a standard anticoagulant therapy for patients who have had a heart attack. In recent years, aspirin therapy has been extended to individuals with a history or at risk for stroke (apoplexy) and phlebitis. It has even been proposed that every person over 50 years of age should take aspirin.

However, some people cannot take aspirin due to allergy, erosion of the stomach lining etc. Furthermore, research has shown that aspirin prevents heart attacks in about 40 percent of patients taking aspirin. Thus, it is desirable to determine which people will respond to aspirin and which will not in order to administer other anticoagulant or antiplatelet medication.

Platlet aggregration is recognized as an important step in the formation of a blockage which will cause a myocardial infarction and unstable angina. Platlet aggregration is based on glycoprotein gpIIb/IIIa. Different forms of this glycoprotein have been known. Weiss et al, Tissue Antigens, 46: 374-381 (1995), Kunicki et al, Molecular Immunology 16: 353-60 (1979). Methods for determining various polymorphisms may be done by DNA analysis. Newman et al, Journal of Clinical Investigation 83:1778-81 (1989). It has been reported that patients having one polymorphic form of the PI gene have a higher incidence for acute coronary thrombosis, particularly in patients younger than 60. Weiss et al, New England Journal of Medicine 334(17):1090-1094 (1996). However, these findings were contradicted by Ridker, et al, Lancet 349: 385-388 (1997) with comments in Lancet on pages 370-371, 1099-1100 and 1100-1. Adding to the debate, it was recently published that platelet aggregation from haplotype PI^{A2} containing individuals are less inhibited by aspirin at certain concentrations than individuals homozygous for haplotype PI^{A1}. Cooke et al, Lancet 351: 1253 (1998).

Resolving the issue for people at risk of heart attacks, stroke and other thrombogenic disorders is desirable, particularly in distinguishing between those who can take aspirin or who should take other medication which is more costly and with greater side effects.

Experimental protocol

Blood samples are taken from 50 healthy individuals ages 50-55. Family history and personal histories of heart disease and other thrombogenic disorders are recorded. White blood cells are collected and genomic DNA is extracted from the white blood cells, PCR amplified and the sequence determined by ASO or sequenced as in the Examples above using different primers and probes. Newman et al., Journal of Clinical Investigation 83:1778-81 (1989). As before, PCR primers and ASO probes are designed to type these individuals for exon 2 to determine which base exists at nucleotide position 1565: a T or a C. at the amino acid level, codon 33 is changed from a leucine to a proline.

Individuals having haplotype PI^{A2} either in homozygous or heterozygous form are instructed to either take high dosages of aspirin (2000 mg per day) or not take aspirin and given other medication appropriate for their individual needs. Individuals homozygous for haplotype PI^{A1} are instructed to take aspirin at low dosages (350 mg per day)

The present invention is not to be limited in scope by the specific embodiments described herein. Indeed, various modifications of the invention in addition to those described herein will become apparent to those skilled in the art from the foregoing description and accompanying figure. Such modifications are intended to fall within the scope of the appended claims.

Various publications are cited herein, the disclosures of which are incorporated by reference in their entireties.